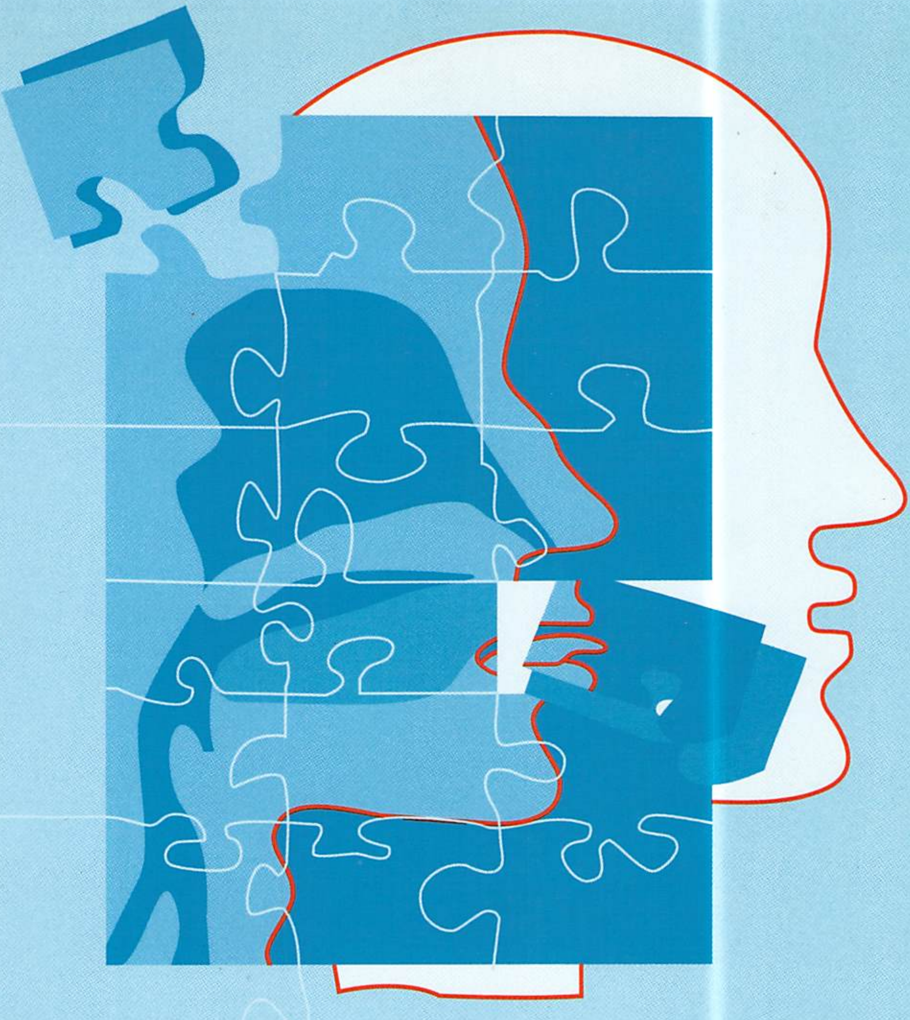


PRAGMATICS & COGNITION

Volume 14, Number 2 (2006)

Special Issue on Cognition and Technology

Distributed Cognition

PRAGMATICS & COGNITION is an interdisciplinary journal seeking to bring together such disciplines as philosophy, linguistics, semiotics, cognitive science, neuroscience, artificial intelligence, ethology, and cognitive anthropology, among others. The journal seeks to explore relations of all sorts between semiotic systems as used by humans, animals and machines, in connection with mental activities: logical and causal dependence; condition of acquisition, development of loss; modeling, simulation of formalization, shared or separate biological and neurological bases; social and cultural variation; aesthetic expression; historical development; etc. Pragmatics & Cognition's basic assumption is that the proper understanding of mental life and inter-personal relations requires an intensive and thoughtful exchange of views across disciplines.

PRAGMATICS & COGNITION pays special attention to the growing interest in the relationship between technological advances and cognition – a field that is intimately related to the journal's basic concerns. Each Special C&T Issue will be co-edited by a Guest Editor and Dr. Itiel Dror: http://www.ecs.soton.ac.uk/~id/technology.html

PRAGMATICS & COGNITION contains original articles, research reports, scholarly notes and book reviews.

PRAGMATICS & COGNITION invites relevant contributions. Authors are advised to consult the Guidelines given on the inside back cover and on www.benjamins.com.

PRAGMATICS & COGNITION subscription information
Annual subscription Volume 14, 2006 (3 issues, approx. 600 pages)
Institutions/Libraries EUR 378.00
Includes postage & handling and on-line access.
Individuals may apply for a special subscription rate of EUR 80.00. Private subscriptions are for personal use only, and must be pre-paid and ordered directly from the publisher.
Individual members of the International Pragmatics Association may order the journal at a special discount.

John Benjamins Publishing Company
Subscriptions
P.O. Box 36224
1020 ME Amsterdam
The Netherlands
Fax: +31 20 6739773
subscription@benjamins.nl

www.benjamins.com

PRAGMATICS & COGNITION is indexed in The Philosophers Index, and PsycINFO.

ISSN 0929-0907
© 2006 John Benjamins Publishing Company, Amsterdam, The Netherlands

Attention before copying:
In the USA: Authorization to photocopy items for internal or personal use is granted by John Benjamins Publishing Company, provided that the appropriate fee is paid directly to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, USA. Fees are subject to change without notice.
In the rest of the world: Permission to photocopy must be obtained from the copyright owner: John Benjamins Publishing Company, P.O. Box 36224, 1020 ME Amsterdam, The Netherlands.

All views or conclusions are those of the authors of the articles and not necessarily those of the editorial staff or the publisher.

PRAGMATICS & COGNITION

VOLUME 14 NUMBER 2 2006

Editor

Marcelo Dascal
Tel Aviv University

Associate Editors

Jens Allwood
University of Göteborg

Itiel E. Dror
(Cognition & Technology)
Southampton University

Benny Shanon
The Hebrew University of Jerusalem

Stephen Stich
Rutgers University

Yorick Wilks
University of Sheffield

Assistant Editors

Edson Françozo
State University of Campinas

Yaron Senderowicz
Tel Aviv University

Editorial Assistant

Dana Riesenfeld
Tel Aviv University

John Benjamins Publishing Company
Amsterdam/Philadelphia

PRAGMATICS & COGNITION

Board of Consulting Editors

Jerome S. Bruner (Psychology)
New York

Andy Clark (Cognitive/Computer Sciences)
Washington University, St. Louis

Umberto Eco (Semiotics/Cognitive Science)
University of Bologna

Barbara Gorayska (Cognition and Technology)
Cambridge, UK

James Higginbotham (Philosophy/Linguistics)
University of Southern California, Los Angeles (Philosophy)

Jaakko Hintikka (Philosophy/Logic)
Boston University

Sachiko Ide (English/Pragmatics)
Japan Women's University, Tokyo

Jan Nuyts (Pragmatics/Cognitive Science)
University of Antwerp

Dan Sperber (Anthropology/Pragmatics)
CREA-Ecole Polytechnique, Paris

Deborah Tannen (Pragmatics)
Georgetown University, Washington D.C.

Daniel Vanderveken (Philosophy)
Université du Québec, Trois Rivières

Boris M. Velichkovsky (Neuroscience/Computer Science)
Dresden University of Technology

Anna Wierzbicka (Linguistics)
Australian National University, Canberra

SPECIAL ISSUE ON COGNITION AND TECHNOLOGY

Distributed Cognition

Edited by
Stevan Harnad and Itiel E. Dror

Table of contents

Call for papers 207

Introduction

Distributed cognition: Cognizing, autonomy and the Turing Test 209
Stevan Harnad and Itiel Dror

Articles

A framework for thinking about distributed cognition 215
Pierre Poirier and Guillaume Chicoisne

Distributed cognition: Domains and dimensions 235
John Sutton

Distributed cognition: A methodological note 249
David Kirsh

Radical changes in cognitive process due to technology: A jaundiced view 263
Arthur M. Glenberg

The grounding and sharing of symbols 275
Angelo Cangelosi

Collaborative tagging as distributed cognition 287
Luc Steels

Thinking in groups 293
Todd M. Gureckis and Robert L. Goldstone

Distributed learning and mutual adaptation 313
Daniel L. Schwartz and Taylor Martin

Distributed cognition, representation, and affordance 333
Jiajie Zhang and Vimla L. Patel

Categorization and technology innovation 343
Jeffrey M. Stibel

Crime scene investigation as distributed cognition 357
Chris Baber, Paul Smith, James Cross, John Hunter, and Richard McMaster

Web Search engines and distributed assessment systems 387
Christophe Heintz

Speech transformation solutions 411
Dimitri Kanevsky, Sara Basson, Alexander Faisman, Leonid Rachevsky, Alex Zlatsin, and Sarah Conrod

Computer-aided translation as a distributed cognitive task 443
Barbara Dragsted

Call for Papers

Fourth special issue in the series *Cognition and Technology*

Learning Technologies and Cognition

Editor: Itiel Dror

Learning technologies have been taking an increasing role in almost all learning environments. They are used in a variety of informal and formal educational environments, from early years to university level and throughout adulthood, as well as in many commercial, industrial, and governmental settings. With the greater use of learning technologies it is critical to better understand how they interact with human cognition. Both in terms of how they may facilitate and enhance (as well as hinder) learning, and also in terms of how they affect the way we learn and acquire information, and the nature of cognition.

These issues pertain to specific technologies and to learning objectives. Specific technologies (and their usage) are important to understand in their own right; for example, how the use of electronic boards and visualization tools, e-learning, synchronic vs. a-synchronic remote learning, blackboard, simulation, virtual realities, and other technological learning environments affect learning and the learner. But also the learning technologies need to be considered and understood in light of learning objectives: not only the acquisition of information, but also the ability to retain and use it and the assessment of the effectiveness of the learning process. When considering how best to use learning technologies (and their vulnerabilities) one needs to be able to determine which learning materials and objectives are best suited for these technologies, which learning tools are most appropriate, and how to best use them. Furthermore, a fundamental issue to address is if and when learning technologies should replace traditional learning and when and how should learning technologies be blended with traditional learning.

Original and high quality papers that examine learning technologies either from an academic or from a practical perspective will be considered

Pragmatics & Cognition 14:2 (2006), 207–208.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

for publication. The first special issue of *Pragmatics & Cognition* devoted to Cognitive Technologies is now going to be published as a book. It is hoped that the Learning Technologies special issue will also appear in book form in the future.

Deadline for submissions: 30 June 2007

Publication: Summer 2008

Send submissions and any inquiries to Itiel E. Dror, id@ecs.soton.ac.uk

Distributed cognition

Cognizing, autonomy and the Turing Test

Stevan Harnad and Itiel Dror
University of Southampton

Some of the papers in this Special Issue distribute cognition between what is going on inside individual cognizers' heads and their outside worlds; others distribute cognition among different individual cognizers. Turing's criterion for cognition was for individual, autonomous input/output capacity. It is not clear that distributed cognition could pass the Turing Test.

Cognition is what cognizers do: What are cognizers? They are autonomous systems with certain I/O (input/output) performance capacities; and cognition is the internal process that generates that I/O capacity.

What is I/O capacity? It is what higher (and possibly lower) vertebrates are able to do in their worlds: In the case of our own species, it is the capacity to pass the Turing Test. Cognition is what organisms' brains do to enable them to do all the things that organisms can do.

And, until further notice, cognition takes place entirely within the brains of cognizers: It starts with the I in the I/O and ends with the O, skin and in, stretching from the proximal projection of distal objects, events and states onto the cognizer's sensory surfaces to the proximal projections of motor patterns onto the cognizer's effector surfaces, but no further (this is sometimes called "narrow cognition"). The causes and effects stretch more distally, but not the cognition; cognition begins and ends at the cognizer's sensor and effector surfaces. (Fighting words, in our ecumenical era of "distributed cognition"!)

So if, to a first approximation, cognizing is whatever goes on inside the brains of organisms to give rise to their performance capacity, then cognition is the functional substrate — the generator — of their know-how. And their performance is that of individual, autonomous systems. It's a solo performance, not like a symphony orchestra, which is in fact many performers. A violinist is a performer; a string quartet is not: it's four performers, performing together.

Pragmatics & Cognition 14:2 (2006), 209–213.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

The violinist knows how to read music; the quartet does not. An organist knows how to play 12 notes at once (and to sing and beat rhythm with any remaining appendages at the same time); a musician may be able to play, successively, a string, wind, keyboard and percussion instrument (possibly even a few of them at once), but no musician can play an orchestra — and the conductor merely leads it, as a shepherd leads his flock; he does not play it. When a conductor is auditioned, it is to test how he directs an ensemble of players, not how he plays the oboe.

There exists an audition for cognition too: The Turing Test (TT). The candidate must be able to exhibit the performance capacities *of* a normal human being, indistinguishably *from* a human being, *to* a human being. But just as it would be regarded as cheating on an exam if the "candidate" consisted of a student plus all his friends and hirelings, sitting the examination jointly, there is a bug in the original version of the TT, which was meant to be conducted by email: If the candidate has performance capacity indistinguishable from that of a real human pen-pal, then it successfully passes the TT. But the purpose of the TT was so that we could design a machine that could pass it, in order that we should at last understand what cognition is, what generates the know-how of an autonomous system like us. A whole team of real human pen-pals could of course also pass the test, pretending to be just one pen-pal, but that would not teach us anything at all about cognition; it would simply be cheating, and the outcome would be trivial and uninformative (except perhaps concerning the social psychology of composite play-acting).

How to close this loop-hole that would trivialize the TT? Don't restrict it to email: Require that the candidate should be a robot we can see as just one individual autonomous system like ourselves. That way we not only eliminate the possibility of collective play-acting, but we can also test the candidate's full sensorimotor I/O capacity to see whether it is indistinguishable from our own. After all, despite the remarkable and undeniable expressive power and universality of natural language, human cognizers are capable of a lot more I/O than just email in and email out.

But what is and isn't a single autonomous system? Certainly a robot controlled remotely by a team of technicians would be neither an autonomous system nor a TT candidate in any interesting sense. An organism is a single autonomous system, but so are its cells or even its organs, in the right support media. So an autonomous system may itself be composed of autonomous systems as parts. And so far we are talking about biological autonomy. The logic of autonomy probably breaks down with engineering and physics, where a plane, a pendulum and a positron are all "autonomous systems" in some sense.

So autonomy is really only interesting in the cognitive/behavioral domain — which brings us back to Turing-scale I/O capacity.[1]

Let us agree, without further hair-splitting, that insofar as the TT is concerned, any sort of *conscious* multi-person collaboration in generating the robot's I/O capacity would be cheating, or begging the question that the TT is meant to answer: How does the brain generate its I/O capacity? It doesn't generate it by recruiting the brains of a team of collaborators. Is it logically or empirically impossible that the TT could be passed by a bona-fide superordinate cognizer emerging from the interactions of many individual cognizers in the same way that a superordinate slime mold emerges from many individual amoebae? It is not logically impossible, but let us admit that nothing even faintly like it is anywhere in sight empirically, outside of science fiction. What do we see, then, that aspires to be called "distributed cognition"?

First, we see the distal objects of the skin-and-in proximal cognition we have already admitted. Is there a "wider cognition", consisting of what is going on in a cognizer's brain, plus the external objects that project their shadows onto the cognizer's sensory surfaces, that are manipulated by the cognizer's effector actions, and that are the ("intentional") objects of the cognizer's thoughts? It seems that if we admit this wider spectrum of objects into cognition, we lose both the autonomy of the I/O system that is being T-tested and the TT itself: How can the I's and O's and their sources and sinks be part of the self-same autonomous system that is being T-tested for its I/O capacity? (Yet this seems to be what Sutton, Kirsh, Schwartz, and Zhang in their respective position papers [in this Special Issue] mean by "distributed cognition".)

The second candidate for "distributed cognition" seems to be social, with the interactions among multiple cognizers being called "distributed cognition". (The Cangelosi, Steels, Goldstone, and possibly the Poirier position papers in this Special Issue seem to be thinking along these lines.) We have agreed that a team of cognizers — collaborating to pretend to be a single pen-pal in the email TT, or remotely controlling a robot in the robotic TT — would be cheating and uninformative. Is there any sense in which multiple cognizers can be said to be doing unitary rather than multiple (additive, or even multiplicative) cognizing? This would seem to call for a superordinate TT test — but testing for what I/O capacity? And compared to what? Individual people are indisputable, paradigmatic cognizers: They are cognizing, if anything is. Their individual I/O capacities are accordingly the gold-standard against which the TT candidate is judged. What would be the corresponding standard for a superordinate TT for distributed cognition? What would the distributed cognizer need to be able to do? And compared to what? And if human I/O capacity, TT and autonomy are set aside, what is left to nonarbitrarily call "cognizing", whether local or distributed? (These questions are merely raised here: They are not necessarily decisive.)

For the remaining position papers — Glenberg's (in this Special Issue) and Harnad's (2005), as well as the five empirical review papers (in this Special Issue) — "distributed cognition" is really just synonymous with "information processing" and indifferent as to whether it is indeed cognizing, or autonomous. Does it matter? Could we just scrap "cognition" and speak about "functioning", whether local or distributed? The reader is left to judge. What is indisputable is that the practical potential of both local and distributed function, involving both humans and machines in the online age, is enormous and just beginning to be exploited.

Note

1. Even here, with distributed cognitive processing possible within an individual cognizer according to the paradigms of neural networks, parallel distributed processing, and connectionism, the nature and boundaries of a system and its parts may be fuzzy or ambiguous (Dror and Dascal 1997). The overall behavior of an individual (arising from distributed processing or from centralized processing), however, is clear, with I/O capacity taken to reflect cognitive processing and capacity (although this view has its challengers, such as Searle 1980).

References

Dror, I.E. and Dascal, M. 1997. "Can Wittgenstein help free the mind from rules? The philosophical foundations of connectionism". In D. Johnson and C. Erneling (eds), *The Future of the Cognitive Revolution*. Oxford: Oxford University Press, 293–305.

Harnad, S..2005. "Distributed processes, distributed cognizers, and collaborative cognition". *Pragmatics & Cognition* 13(3): 501–514.

Searle, J.R. 1980. "Minds, brains, and programs". *Behavioral and Brain Sciences* 3: 417–424.

Turing, A.M.. 1950. "Computing machinery and intelligence". *Mind* 49:433–460.

About the authors

Stevan Harnad, born in Hungary, did his undergraduate work at McGill and his doctorate at Princeton and is currently Canada Research Chair in Cognitive Science at the University of Quebec, Montreal and adjunct Professor at Southampton University, UK. His research is on categorisation, communication and cognition. Founder and Editor of *Behavioral and Brain Sciences*, *Psycoloquy*, and *CogPrints Archive*, he is past President of the Society for

Philosophy and Psychology, Corresponding Member of the Hungarian Academy of Science, and author and contributor to over 150 publications: http://www.ecs.soton.ac.uk/~harnad/

Itiel Dror is a Senior Lecturer in cognitive sciences at the University of Southampton, UK. He holds a number of graduate degrees, including a Ph.D. in cognitive psychology from Harvard (USA). He specializes in the fields of human cognition and behaviour, training and skill acquisition, technology and cognition, expertise, and biometric identification. Dr. Dror has worked in universities around the world and conducted research and consultancy to numerous organizations, including the UK Passport Services, the USA Air Force, the Japanese Advance Science Project, the European Aerospace Research & Development Agency, the Israeli Aerospace Industry, the BBC, and many commercial companies. He has published widely in both basic science and applied domains. For more information: http://www.ecs.soton.ac.uk/~id/

A framework for thinking about distributed cognition

Pierre Poirier and Guillaume Chicoisne
Department of philosophy and Cognitive Sciences Institute / Cognitive Sciences Institute, Université du Québec à Montréal

As is often the case when scientific or engineering fields emerge, new concepts are forged or old ones are adapted. When this happens, various arguments rage over what ultimately turns out to be conceptual misunderstandings. At that critical time, there is a need for an explicit reflection on the meaning of the concepts that define the field. In this position paper, we aim to provide a reasoned framework in which to think about various issues in the field of distributed cognition. We argue that both relevant concepts, distribution and cognition, must be understood as continuous. As it is used in the context of distributed cognition, the concept of distribution is essentially fuzzy, and we will link it to the notion of emergence of system-level properties. The concept of cognition must also be seen as fuzzy, but for a different reason: due to its origin as an anthropocentric concept, no one has a clear handle on its meaning in a distributed setting. As the proposed framework forms a space, we then explore its geography and (re)visit famous landmarks.

Keywords: distributed cognition, multiagent system, swarm intelligence, emergence, bounded cognition, sociality

1. Introduction

Are companies, organisations and communities in the cognitive business just as we are? If you do your writing on OpenOffice's text editor, you are using an artifact for the conception of which a loosely connected community of individuals, spread throughout the world, have collaborated towards the common goal of providing free, open-source software to individuals worldwide. Can that community be said to be, in some sense, a collective cognitive system, performing distributed cognition, whose collective behaviour was the

Pragmatics & Cognition 14:2 (2006), **215–234**.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

construction of an artifact no individual human programmer could ever construct by himself? Perhaps. But consider this: How is what they do different, really, from the process that constructed MS-WORD? Apart from the surely contingent fact that Microsoft's programmers and engineers were hired and paid by the corporation, is there any relevant cognitive difference between the community of people that Microsoft hires to write its product and the community that selflessly collaborate to produce OpenSource software? Is Microsoft a cognitive system? What about Ford or Toyota? And what about universities, news agencies, courts of law, parliaments, and military units?

Take a court of law. Its purpose is explicitly cognitive: to determine, beyond reasonable doubt, the truth-value of certain propositions (X is guilty of crime C). To do so, a number of individuals have specific roles to play, each of them cognitive. Police officers have to *investigate* and *find facts relevant to the case*; defence attorneys have to *research* (the *declarative external memories* that are) the various codes (criminal, civil, etc.) and precedents repertories looking for laws and precedents that are relevant to their case and, based on those, both *interpret* and present the case in a manner that is most favourable to the accused and *rebut* the prosecution's arguments; prosecutors have to do the same but in a manner that is least favourable to the accused; the judge has to *ensure the validity of the whole procedure*; and finally the jury has to *listen* to all the evidence presented to it, *evaluate* it, *deliberate*, and *render judgment* as to the truth or falsity of the propositions presented to it by the court. The system is so built that, when everything works well (which is not always the case), none of the beliefs of the individuals involved *determines* the outcome. The decision process is, in some sense, supra-individual. As a cognitive process, evaluating the truth of the propositions is a system-level affair. This is the difference between living in a society with a justice system versus living in a society with vigilante justice (in French: the difference between "un système de justice" and "un justicier").

Now, is this description just a loose and ultimately unproductive analogy with the human mind? Perhaps courts of law were designed with an eye on the only model of a truth-finding system that was at hand at the time — the human mind. Or perhaps, historians say, the human mind was conceived (in its modern form) with an eye on the only truth-finding system known at the time — courts of law. Either proposition would explain the analogy. But the question remains: Is this mere analogy or is there more substance to the claim that there can be supra-individual systems to which one can attribute 'distributed cognition'? Some are quite content to view companies as cognitive systems of sorts, possessing institutional memory (which, like ours, may be tacit or explicit, declarative or procedural), decision capacities, creativity, and so on (Nonaka and Tekeuchi 1995; Hutchins 1996). Is this just conceptual confusion or clever marketing hype?

The foregoing only focused on humans. But what about ant colonies or other collectives of eusocial animals? It is possible to view them as super-organisms, that is, organisms made up of other organisms; or as Minsky (1985: 165) described them: "Genetically, the swarms of social ants and bees are really multibodied individuals whose different organs move around freely". And if this can be said from a physical standpoint, shouldn't cognitive scientists be ready to study the (distributed) cognitive properties of super-organisms? Some are already eager to speak of "swarm intelligence", "ant colony optimization", and the like (Bonabeau et al. 1999, Dorigo and Stützle 2004). And what about the multi-agent systems built by engineers, which will come to play an increasingly important role in a high-bandwidth networked society? Are these people just confused or are they on to something?

To address these questions, two conceptual options are possible. The first, conservative, insists on the differences between supra-individual and individual human cognition, thus enforcing a traditional and strict reading of the relevant concepts. The second, liberal, emphasizes the similarities between supra-individual and individual human cognition, thus opening the possibility of a reorganization of the conceptual landscape of cognition. Both options, we believe, illegitimately impose sharp readings on two fuzzy[1] concepts: distribution and cognition. In the context of cognition, distribution, as we'll argue, is essentially fuzzy; and in the context of distribution, cognition, as we'll also argue, must, at least for now, be given a family-resemblance definition, best captured by a multidimensional fuzzy concept. In this paper, we propose a framework that restores the essentially vague nature of the relevant concepts and explores how various candidates for the title of supra-individual cognizers (e.g., neural networks, ant colonies, multi-agent systems, human organizations) fare in this new conceptual landscape.

2. A notion of distributed cognition

Proponents and opponents of distributed cognition have a definite, though usually not clearly explicit, idea of what "distribution" means in the context of cognition. We believe this underlying idea is linked to the notion of emergence. Truly distributed cognition is emergent cognition. One truly has a distributed cognitive system when one has a system where a new cognitive property *emerges* from the interaction between the system's components, which may themselves be cognitive systems. This link between distributed cognition and emergence is fine, we believe. The problem is that people usually work with an all or nothing conception of emergence inherited from the early 20th century. Emergentism is often caricatured as the thesis that *the whole is more than the sum of its parts*. Note that something either is or isn't more than the sum of its parts. *This* concept of emergence is binary (in logic: classical). Hence, viewing distributed cognition as emergent cognition, something either is or isn't a case of distributed cognition. But things need not be this way. We will ground this discussion on a notion of emergence (inspired by engineering) that makes emergence an essentially fuzzy concept.

W. Wimsatt (1986), a philosopher of biology trained as an engineer, offers a notion of emergence that is perfectly consistent with the current ontology of science (as opposed to the form of emergentism that was popular, say, at the beginning of the 20th century). He defines emergence as a failure of aggregativity. Take a property P of a system S, s_1 to s_m being the *m* components of S, p_1 to p_n the *n* properties, $p_1(s_1), p_2(s_1), \ldots, p_n(s_1), p_1(s_2), p_2(s_2), \ldots, p_n(s_2), \ldots p_1(s_m), p_2(s_m), \ldots, p_n(s_m)$ properties of *S's* components, and the organisation or interaction mode *F* of these component properties. P of S may be defined thus:

$$P(S) = F[p_i(s_j) \text{ for i} = 1 \text{ to n and j} = 1 \text{ to m}]$$

P(S) is aggregative to the extent that four conditions are respected:

1. *Condition IS (invariance under substitution):* $P(S)$ is invariant under intersubstitution of the parts of S with one another or under substitution of one or more of the parts with other parts from a domain of relevantly similar parts.
2. *Condition QS (qualitative invariance)*: $P(S)$ remains qualitatively similar (differing only in value) under addition or subtraction of parts.
3. *Condition RA (re-aggregation):* The composition function for $P(S)$ is invariant under operations involving decomposition or re-aggregation of parts.
4. *Condition CI (cooperation/inhibition)*: There are no cooperative or inhibitory interactions among the parts of the system.

Take a pile of books. It is (in some sense) a system and has a number of properties, among them a certain mass. Consider the property *Mass*(pile of books), or for short M(b). Now, you can interchange the position of the books in the pile or replace one book by one that is similarly relevant with respect to M(b), that is replace it with another book of the same mass. Clearly M(b) is invariant under such substitutions: condition IS is satisfied. Also, books can be added or subtracted from the pile without the pile either losing its property of mass or mass becoming some qualitatively different property. M(b) is qualitatively invariant under the operations of addition and subtraction of books: condition QS is satisfied. Moreover, individual books from the pile may be decomposed (e.g., put trough a shredder) or re-aggregated (good luck!) and M(b) will stay the same. M(b) is invariant under the operations of decomposition or re-aggregation: condition RA is satisfied. Finally, with respect to M(b) there are no relevant interactions (cooperative or inhibitory) between the books in the pile: condition CI is satisfied. It follows that M(b) is a completely aggregative (hence non-emergent) property. The height of the pile, H(b), on the other hand, fails condition CI, for there are inhibitory interactions between the books in the pile so that different ways of organising the books (vertically, horizontally) yield different heights. Since H(b) does not fail any other conditions of aggregativity, we may say that H(b) is a largely aggregative property (or a mildly emergent one).

As Wimsatt points out, viewing emergence as the failure of conditions of aggregativity has the advantage of seeing emergence as a continuum between fully aggregative properties and fully emergent properties, both of which are rare in nature. Most properties fall somewhere along this continuum, as determined by the extent to which they satisfy or fail aggregativity conditions 1 through 4. Note also, as Wimsatt points out, that since emergence, thus defined, presupposes the existence of a composition function (*F*), it is not the opposite of reducibility. All systemic properties on the aggregativity/emergence continuum are in principle reducible to the properties of the system's components, their mode of organisation and their interactions. This is what is called "objective emergence": emergent properties are real properties, in principle reducible to lower level properties. Whether they can in practice be reduced depends on our knowledge of (1) the system's components and their properties and (2) the systemic interactions between them. If we know nothing or very little of either or if these are complex (relatively to human's limited computation abilities), then the more a systemic property is emergent the more it will seem to "spring" (almost magically) from nothing. This condition is called "subjective emergence". The important point about subjective emergence is that it is just that: subjective (i.e., an epistemic reflection of our ignorance).

Now that we have explained what we believe the concept of emergence underlying any theoretical reflection about distributed cognition should be, we must do the same with the concept of cognition. Why is this important? Doesn't everyone know what cognition is? In the present context, we need to be careful about that. The novel, but controversial, idea behind the notion of distributed cognition is that we may ascribe cognitive properties to types of systems that haven't been traditionally seen as the bearers of such properties: ant colonies, organisations, research labs, and so on. In this context, a number of things about the concept of cognition should be kept in mind.

(1) We need to define cognition in a way that is not chauvinistic, for instance by tying cognition to human manifestations of cognition. Cognitive systems may need sensors, but they do not need ours (including their felt qualities). Cognitive systems may live in an environment, but they do not need our physical environment, even less our own ecological niche, in order to survive. Cognitive systems may need warning signals that tell them something is going on that threatens their integrity, but they do not need our warning systems (pain, anxiety, fear, stress). Cognitive systems need to persist for some time, but they do not need to do so by being part of biological life on earth. In short, we need a conception of cognition that is not anthropocentric and that allows the ascription of cognitive properties to systems that are not typically viewed as such. Otherwise, the whole issue of whether distributed cognition is possible or not will be decided by conceptual decree, that is, by insisting on an anthropocentric concept of cognition that *makes* distributed cognition impossible. Once a debate over ideas has been reduced to that level, it usually becomes an unproductive fight over the meaning of words.

(2) But the problem that concerns us goes deeper than advice for caution against anthropocentric conceptions of cognition could solve. We have to be open to the possibility that the cognitive property in question may be quite alien to us and that, as a consequence, we may be literally blind to it. Think of it this way. Autists are, it is said (Baron-Cohen 1997), blind to some mental properties in others, that is, they are be mind-blind. The question is: if some truly novel cognitive property emerges out of the interaction of agents, will we be as blind to it as autists are (according to Baron-Cohen) blind to our run-of-the-mill mind? This is related to the classic sci-fi plot-maker: Will we recognize an alien intelligence staring at us right in the face? We don't propose to resolve this old chestnut here (it may actually be impossible to do so). But it is important to keep a broad mind when thinking about cognition in the present context.

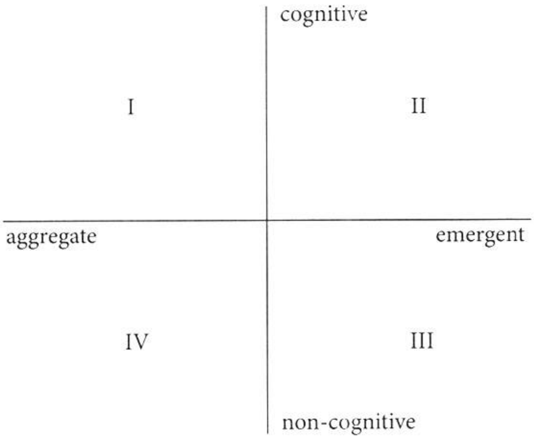
(3) However, the very fact that what we are looking for may be quite alien to us should impose a measure of caution. No one wants to ascribe cognitive properties to Gaia, Jupiter's Red Spot or auto-catalytic chemical reactions. Talk of cognition in the context of distributed cognition invites analogy, metaphor, and loose-talk. While these may be necessary in science, especially when disciplines are young, as any cognitive scientist who ever bought the "computer metaphor" should know, they are also dangerous tools that can quickly discredit any aspiring young field (think of "memetics").

(4) Furthermore, cognitive properties should not be bundled too closely, otherwise illegitimate inferences might be unavoidable. No one wants to infer that ant colonies must be *conscious* because they can collectively optimize some parameter. Or, conversely, no one wants to deny ant colonies some cognitive properties because they are obviously not conscious. Bearing these points in mind, we propose to explicitly view cognition as a cluster concept, that is, a concept that denotes a loosely tied family of properties. The cluster we propose is not meant as a definition of cognition, but as a *diagnostic tool* for picking out systems that may suitably be viewed as cognitive.[2]

P(S) contributes to S's cognitive abilities to the extent that:

- *Condition AD (adaptability)*: P is involved in S's capacity to adapt its behaviour to match changing environments;
- *Condition IP (information processing)*: P is involved in S's processing information from its environment;
- *Condition I (intentionality)*: P makes structures in S about structures in its environment;
- *Condition C (consciousness)*: P makes S, or structures in S, conscious (in some of the various senses of the term).

These conditions are not disjoint, and they are not mutually exclusive. To insist, they are meant as diagnostic criteria such as one finds in, e.g., the Diagnostic and Statistical Manual (DSM) used by psychiatrists to diagnose mental illness. If a system has none of these properties, it will be said to be non-cognitive; if it has a few, it will be said to be minimally cognitive; if it has all four, it will be said to be highly cognitive. Given both the notions of an emergent property and



that of a cognitive property, we will define the notion of a *distributed cognitive property* as any system property that is both cognitive and emergent. Note that since both notions are matters of degree, and since both are independent, they form a space where various properties of systems may be plotted. We will refer to that space as the emergence/cognition space (E/C space). Temperature in a gas will fall squarely in quadrant III of the E/C space, whereas the mass of the gas will fall in quadrant IV.

In the remainder of this paper, we will explore various types of systems to see where they fall in the C/E space. Distributed cognitive systems, if there is such a thing, will be shown to fall in quadrant II.

3. A benign form of distributed cognition

The previous section presented quite abstract ontological and conceptual stuff. The time has come to make things much more concrete by applying our definitions to specific cases. We start by a case that, we believe, is an uncontroversial case of distributed cognition (except that it is not the type of cases people usually have in mind when they talk about distributed cognition — we'll get to those in due time). The function of this section is to test our definition and set out a clear case of distributed (*qua* emergent) cognition before we address the more fuzzy cases.

Take the case of a formal neural network. It is made up of two types of (formal) entities: neurons (N) and connections (C) — capitals indicate types of entities. In standard models (such as multi layer perceptrons), neurons have a number of properties: an activation state (as), an integration function (if), and a transfer function(tf). To use the above formalism, N(as, if, th). They may have other properties, but let's stick to those to keep things simple. Connections, in turn, have a connection weight (w) and a connectivity <ni, nj>: C(w, <ni,nj>). Any specific neural network NNet of a large class of neural networks may be defined with only this simple set of entities and properties. Notice that none of NNet's components' properties deserve to be called "cognitive". Perhaps a component's integration and transfer functions can be called "computational" (a kind of analog computation), but this is not important here. Now, as everyone knows, systems such as NNet have been trained[3] to manifest various cognitive properties. Let's say NNet has been trained to categorize objects (for instance, faces). A qualitatively different type of property, CAT (for categorization) has emerged from the interaction of N's components. No hocus pocus here, as anyone can write the equation describing the emergent property of CAT, for an *n* neurons network:

$$\mathrm{CAT}(N) = F[as(n_i), \mathrm{if}(n_i), \mathrm{tf}(n_i); \text{for } \mathrm{i} = 1 \text{ to } \mathrm{n}; \mathrm{w_k}, <\mathrm{n_i}, \mathrm{n_j}> (\mathrm{c_k}); \text{for } (\mathrm{i}, \mathrm{j}) = (1,2) \text{ to } (\mathrm{n} - 1, \mathrm{n}) \text{ with } \mathrm{i} < \mathrm{j} \text{ and } \mathrm{k} = 1 \text{ to } \mathrm{n}(\mathrm{n} - 1)/2]$$

If, given its connectivity, NNet is a multi-layer perceptron,[4] then *F* is simple linear algebra (vector by matrix multiplication, summation, etc.). Of course, psychologists who use neural networks to model and study categorization are not interested in this equation. They're interested in the global properties of the system. Does the net categorize this set of data, better than this one? Does its categorization ability generalize well? How is its ability influenced by time? To ask and answer all of these questions, they need vocabulary and measurements that apply to the global behaviour of the network. That is, they have a genuine epistemic project (understanding categorization), which can only be pursued at the level of the global system.

Where does S's categorization capacity fall on the C/E space? Before it was trained to manifest CAT, N behaved inadequately in its environment[5] (as measured by, e.g., RMS error). CAT thus contributes to N being adapted to its environment (low RMS error). Depending on the neural network involved, learning may be ongoing throughout ontogeny, in which case CAT will change over time to reflect the changing categorical structure or the environment, provided it does change. CAT also makes NNet sensitive to the categorical structure of faces in the environment. Given CAT, NNet can now behave in a manner that is sensitive to that structure. Moreover, NNet has CAT because its hidden layer activation space has been partitioned in such a way that sub-spaces are correlated with particular faces in the environment. Whether this correlation is sufficient for full-blown intentionality, or whether it is only a minimal form of intentionality is a matter of debate. Nevertheless, CAT does satisfy condition I to some extent. Finally, CAT is not part of NNet's (or NNet's states) consciousness, since NNet is certainly not conscious.

How about emergence? CAT is a typical emergent property. It fails condition IS: CAT is invariant when neurons are intersubstituted (neurons in ANN are usually clones) but certainly not when connections are. This is all very dependent on the exact nature of the ANN used, but in some cases a few connections may be inter-substituted with only a slight decay in performance.[6] Hence CAT mostly (though not totally) fails condition IS with respect to its trained connections. CAT also fails conditions QS, RA and CI (for lack of space, we leave it to the reader to convince himself of this fact). It follows that CAT falls in Quadrant II, although it does not stand at its extreme (coordinate 1,1).

Categorization in a neural network is a case of a *distributed cognitive property*. Of course, categorization in neural networks, as a case of distributed cognition, is unlikely to generate any debate. The difficult and problematic cases of distributed cognition are those where the system's components themselves have cognitive properties or are themselves agents. But before we turn to full-blown agents, and to get a feel for the framework, let's take a case that is a little more challenging, that of swarm intelligence.

4. Distributed cognition between dummies: Swarm intelligence (SI)

Distributed cognition in neural networks is uncontroversial. The case for distributed cognition becomes contentious when cognitive properties are said to emerge from the collective behaviour of agents. But the situation, we'll argue, is exactly parallel to the previous one; hence, it should not raise any eyebrow. Let's start with simple agents, dummies, before we move on to the really controversial case of interacting agents with cognitive properties.

Swarm intelligence (SI) is the name given to the global behaviour of large groups of locally interacting artificial agents or animals, such as bees building their hive, ants foraging for food, or birds and fish flocking. The components of SI systems are agents, which are endowed with capacities (or properties). These agents and their properties, together with the way they interact given their properties, form the "agent level" of the SI system. The keyword in SI is the word *simple*. Agents are usually simple (simple rules, simple representations) and undifferentiated (though there may be a small number of categories of agents). As for most agents considered in the field of Multi-Agent Systems, their perception capacities are local: perception of their neighbour's behaviour or of traces left in the environment. Their action capacities are also local: moving, leaving traces in the environment, holding pieces of food. These local capacities completely determine their means of interaction with each other and with their environment. Yet, despite the purely local attributes of components at the agent level, global properties such as coordination emerge at the system level.

Consider a classical example of SI: ant foraging. All ants in an ant colony (AC) manifest three simple behaviours when they forage for food: (1) when no pheromone trail is present nearby, they move randomly; (2) when a pheromone trail is present nearby, they move toward it; (3) when a piece of food is present nearby, they pick it up and bring it back to the nest while leaving a pheromone trail. Now consider a single ant belonging to the whole system engaged in foraging. The ant will move about randomly (behaviour 1) until it finds either a piece of food or a trail of pheromone. If it finds a piece of food, it will bring it back to the nest (behaviour 3) creating a pheromone trace between the food source and the nest. If, instead, it first finds a trail of pheromone, following it (behaviour 2) will lead it to a piece of food which it will bring back to the nest (behavour 3), *reinforcing the trace of pheromone that led it to the food source.* As pheromones vanish from the environment after a while, only the trails that are positively reinforced by many ants will remain in the environment, that is, only the paths that link food sources to nest will remain. Moreover, if numerous sources of food are present in the vicinity of the nest, this phenomenon will particulary favour the closer food sources. From the collective behaviours of ants in the colony a global capacity has emerged: the capacity to *Find the path to the Closest Food source*, call it FCF(AC).

Is this global capacity of the ant colony a case of distributed cognition? FCF allows AC to adapt to its changing environment. As the first found food source is depleted, the next closest source will be found and exploited, and so on. If a new food source "appears" in the environment (e.g., a cookie crumb is dropped by a human child), the same process will bring AC to preferentially exploit it if it is closest to the nest. Does FCF make AC sensitive to information in its environment, and does it allow AC to process that information? In a sense, yes. FCF makes AC sensitive to its environment's food geography (or food distance metric). AC, given FCF, "knows" where the food sources are and what distance they stand from the nest. This nest-centric knowledge is operative for the closest food source, but it is potential for all food sources. FCF is about the closest food source, in the minimal sense CAT above was about faces. Of course, like CAT, FCF is not part of AC's consciousness (for the same reason: AC is not conscious). Although we will not write the equation like we did above, it should be clear that this global capacity is emergent, just as the categorization capacity of neural networks is. FCF is invariant under inter-substitution of individual ants (which, like artificial neurons, are clones) and of pheromones (both individually and positionally). Condition IS is thus satisfied. But condition QS is not, since the number of ants in the colony cannot be reduced beyond a certain point (which is a function of the rate of depletion of the pheromone trace and speed of the ants). Conditions RA and CI fail to some extent too. To explain the global behaviour of the ant colony, we need to use concepts such as "short path" and "distance", but none of these are necessary to describe the behaviours of single ants. It follows from all of this that FCF is a distributed cognitive property that is about as emergent as CAT but less cognitive.

Let's take another example, where humans are the dummies. Consider a campus with no paved footpaths. While the architects could have designed the layout of the paths through the lawn, it has been decided to wait for this layout to "come out" and then to pave it. As students go from one building to another, they leave behind them not a trail of pheromone, but of crushed grass. If numerous students walk the same path, the grass begins to die out, the path becomes apparent and attracts more and more students, with the same positive reinforcement phenomenon seen in the ant example (Brassac and Pesty 1996).

Is this a case of distributed cognition? Consider the systemic function *Design Footpath Layout* (DFL). DFL respects condition AD: if a cafeteria is closed or classrooms change place, DFL will display this new organisation of the environment. DFL mildly respects IP: information about the localization of buildings of interest to students and the connections between them is processed. It also respects I: DFL reflects the spatio-temporal structure of student's agenda. Finally, like FCF, it fails condition C. As for the position of DFL on the aggregativity-emergence of that function, it satisfies condition IS: like individual ants, students are interchangeable (with respect to DFL, of course!). DFL also satisfies conditions RA and CI to some extent. But fails condition QS: the dynamics of the system radically changes when the number of students involved falls below a certain threshold. So, from the standpoint of distributed cognition, DFL falls in the same category as the Find the path to the Closest Food source (FCF) property seen earlier. Note that the emergence/cognitive signature of the two systems is quite similar, which is why they are considered as two instances of the same type of systems, SI systems.

The rationale for the introduction of the second example is to mark the difference between all the capacities that the agents have and those that they are actually using (Brassac and Pesty 1996). While no one doubts that university students have more cognitive capacities than ants, the same type of systemic functions can be obtained through two different implementations: the one with minimally cognitive agents in one situation (ants foraging) and the other with full-blown human cognitive agents in the other (students walking). The important point is that they actually play the same *role* in the system and thus in the emergence of the cognitive systemic property. When full-blown cognitive agents participate in SI systems, it is important to note the different cognitive properties manifested at the agent and the system level. When they participate in the SI system, students think about their next class, perceive their local environment (usually only a small portion of the campus ground), and locally decide where to make the next few steps. They are not thinking about designing footpaths or optimizing lawn to paved ground ratios. They're not aware of any of these system level SI properties. Nevertheless, just as in an ant colony, system level properties do emerge from their local behaviours, driven by their own individual desires and choices.

One might complain that such a process is resource consuming, compared to systems that manipulate explicit representations of the world (be they centrally controlled or decentralized) and where there are coordination processes and direct communication between agents. That is, one might think that explicitly increasing the agent's level of cognitive activity would be more efficient to achieve such goals (find the shortest route to a food source, or the best layout for the footpaths on a campus) than expecting them to be the emergent cognitive systemic property of a group of interacting unintelligent agents. There is no definite answer to this legitimate concern, as any answer would be strongly context dependent. An important point to note, however, is that simply increasing the cognitive capacities of the component agents does not guarantee an increase in the cognitive capacities of the system as a whole, since those are emergent and not aggregative.[7] Specifically, while agents with a high level of cognitive capacities spend time *computing* an optimal solution and *solving various* problems they wouldn't have if they were simpler (conflicts, synchronisation, etc.), dummy agents are already on the job. In some situations, the cost of computing is higher than the cost of acting.[8] To complete the picture, regarding agent's cognitive properties and systemic cognitive properties, there are cognitive behaviours that are not emergent (mostly aggregative from the system's cognitive subparts) and there are emergent behaviours that are not cognitive (like traffic jams).

The important point in swarm intelligence is that global properties emerge without the need of explicit description of the global system at the agent level. As was the case previously with neural networks, the concepts needed to explain the behaviour of the SI system are not needed for explaining behaviour at the agent level, even in the case of global properties that emerge from locally behaving humans (as in the lawn case). In the case of ants, the notion of a shorter path is beyond the analytical capacities of such agents in a reactive architecture. In swarm intelligence, agents usually co-act: they do their local job from their own local agent's perspective. In multi-agent systems, composed of higher-level agents, like a pack of wolfs, agents might collaborate: part of their (inter-)actions is not directly related to the task but to synchronisation or to resolution of conflicts. Somehow, collaboration needs some knowledge that there is a higher level task that is to be accomplished at a systemic level while this knowledge is not present for co-acting, like it happens in swarm intelligence. This strength of SI systems is also its most serious drawback: how do you design agents and environments to get (or, as Varela would say, to enact) the desired emergent property? Two options can be considered: training the system (by adapting agent's models and organisation, as in artificial neural networks), or increasing the complexity of their model in order to fit cognitive properties in. Train them, and you loose the "no-differentiability" constraint which in turn leads to the loss of robustness (as ants will have different role and status, taking one away is risky for the systemic property). Add more cognitive properties to the agents and, not only the latter risk is still present, but also, as we pointed out earlier, you end up with entirely new (emergent!) problems like coordination or conflict resolution. As such, SI systems are a transitional point between the "communities" of neurons that make up an artificial neural network and "communities" of cognitive agents enacting distributed cognition.

5. Distributed cognition between cognitive agents: Multi-agent systems (MAS)

The previous examples (neural networks and swarm intelligence) show that cognitive systemic properties may emerge from the interaction of non-cognitive (neurons) or minimally cognitive (ants) components of systems. The footpath example, in turn, shows that, when the cognitive capacities contributed by the cognitive components of SI systems are separated from the rest of their cognitive capacities, humans may (unwittingly) contribute to the emergence of systemic cognitive properties. All these cases rate high on the emergence dimension but rather low on the cognitive dimension.[9] However, these systemic forms of cognition lay the groundwork for various system-level models of cognitive agents, such as Minsky's Society of Mind (Minsky 1985), which follow a principle sometimes called "the recursion principle" in the MAS community: at a higher level of abstraction, a multi-agent system can itself be viewed as an agent. As we'll see, these models do not rate as high on the emergence dimension, taking sometimes a more aggregative form, such as in a network of experts, each "stupidly" dedicated to a given task in a given context. The time has now come to turn to these systems. Our focus example will be the pack of wolves.

When the pack is hunting, each wolf behaves in a way that gives the pack the behavioural capacity to surround its prey (SUR). No wolf in the pack has that capacity: each wolf can prevent the prey from going off in one direction, and "surrounding" is just a matter of locking every direction, which only the pack can do. With respect to emergence, SUR strongly fails two of the conditions, QS and CI (Qualitative Invariance and Cooperation/Inhibition), as it is impossible to surround a prey with one wolf (QS failure) and as the spatial organisation of the wolves greatly impacts this property. SUR is thus undoubtedly, while not completely, emergent. With respect to cognition, SUR validates to a different extent three of the four criteria, failing only on the consciousness one (C): it clearly is adaptable[10] (condition AD) and it distributively generates actions for the system's subparts (wolves) according to a distributed perception (condition IP). The situation is less clear about condition I (Intentionality) but the exact extent to which this condition is respected will not impair the fact that SUR is a cognitive property of the pack. Surrounding a prey is thus an example of distributed cognition in a group of cognitive agents. The pack of wolves is an entity that has more fangs and claws than the lone wolf, but these are pure quantitative changes, which it an aggregative physical property. It also has, however, the behavioural capacity to surround its prey, which is a qualitatively new property, giving the pack an emergent cognitive property. Nevertheless, although the wolves contribute all their cognitive capacities to the emergent behaviour of the pack, unlike human pedestrians designing footpaths, the wolf pack still shares with previous cases of distributed cognition the fact that it satisfies condition IS (inter-substitution): in a hunting pack, wolfs with similar cognitive and behavioural capacities are interchangeable. What if this last aggregativity condition fails, that is, what if agents are highly heterogeneous, like a group of humans, computers and various supporting devices? And what if it is humans who contribute their full cognitive capacities to the system's global behaviour? To conclude this paper, we propose to address this last question.

Recall that, according to our definition of distributed cognition, true distributed cognition will occupy Quadrant II. Emergent cognitive properties are the sign of distributed cognition. Quadrant I, in which system level cognitive properties are mostly aggregative, is where, on this proposal, we find what Harnad (2005) calls collaborative cognition. In other words, distributed cognition is grounded on qualitative changes, while collaborative cognition is grounded on quantitative ones.

To get a feel for the distinction, imagine five persons who decide to write a book together on a given subject S. The cognitive or intellectual project they set themselves to accomplish can be described as "Writing a book on S". In the extreme case, they might decide that the book is to be made up of five separate and distinct parts and decide that each member of the group is to write one part on her own, without any input from the other authors. The resulting book, they decide, will simply be the sum of parts I through V. In this extreme case the resulting behaviour is almost entirely aggregative (almost as much as the weight of the pile of books). When it comes to collaborative cognition, this is an extreme case, almost to the point of non-collaboration. Of course, the authors still needed to collaborate to the extent of deciding how to divide up the work, which means that collaborative cognition is not purely aggregative (as Wimsatt (1986) points out, almost nothing in nature is). No one (one hopes!) collaborates in this fashion. Usually, each author will read, comment upon, re-write, etc., what the others have written. They will discuss points, debate and perhaps even argue on the book's specific content. Is that a case of distributed cognition? No, we may suppose, because, in spite of all the cooperation and inhibition going on (failure of condition CI), there is no failure of condition QS. If one author falls sick before completion of the project, the others can manage to complete the book. Or someone else may be brought in the project to replace the sick author (respect of condition IS). The book may not be as good (or it may be a better one!), but the authors may still manage *to write a book on S*. Nevertheless, all this collaboration/inhibition is bringing us closer to the fuzzy line between collaborative cognition and distributed cognition. Imagine now, perhaps *per impossibile*, that each author has specific qualities such that no one else could replace her (failure of condition IS). If any one of them falls sick, then no book on S can be written (not now, not ever). The property "write a book on S" in this case has crossed the fuzzy line between aggregativity and emergence. By our definition, we will have to qualify the collective behaviour of the authors as distributed cognition (Quadrant II).

6. Concluding remarks

A few points are worth mentioning here and we propose to address them in concluding. Note that the emergent property is a very specific one: "writing a book on S". And in the extreme condition we presented above (each author has particular qualities that cause the failure of condition IS), that specific emergent property may be seen to be the result of distributed cognition. But can general cognitive properties (such as the capacity to categorize, perceive the environment, or recall something) that are qualitatively new emerge from human cognitive interaction? Before we address this question, let's explore the geography of Quadrants I and II. It can be noted that cognitive systems cluster in specific regions of the Quadrants. We argued that the cognitive properties of SI systems and current neural networks are quite emergent but do not score high as cognitive properties. They cluster to the far right a little above the abscissa (x-Axis). The cognitive properties of brains, we may suppose, score high on both dimensions, with the cognitive properties of human brains scoring highest in the cognitive dimension. Finally, we saw in the previous section that the system-level cognitive properties of groups of humans cognitively interacting score high on the cognitive dimension but are mostly aggregative, with perhaps some very specific properties (e.g., writing books on S) crossing over the into emergence's territory. These properties thus cluster left of center, high above the abscissa. To reframe the previous question: Can human interaction generate cognitive properties that stand close to the (1, 1) coordinate?

We suspect these will be hard to come by. There are a few reasons for this. The first is simple. Humans already manifest many cognitive properties which are diverse and can be quite sophisticated (such as those required to pass the Turing Test). Accordingly, if, at the agent's level, humans make use of the full range of their cognitive properties, few system-level cognitive properties that are likely to be seen are really new. Distribution as emergence sets the bar very high for what will count as distributed cognition in the human case: a qualitatively different cognitive property must emerge from the components' interactions. If no qualitatively different cognitive property can be found, then Ockham's parsimony principle urges us to err on the side of caution and refrain from adding new properties to our ontology where none is needed to account for the world.[11] However, remember that this is a fuzzy bar, because the notion of emergence (*qua* failure of aggregativity) is itself fuzzy. In cases that fall in the region where both concepts apply (like the 35 years old person which is a member (more or less) of both sets, "young" and "old"), we believe *epistemic pragmatism* should rule, that is, counting the case as one of collaborative or distributed cognition rests in the final analysis on the pragmatics of explanation and understanding.

The second reason is linked to the concept of subjective emergence we introduced above. Imagine one truly qualitatively new cognitive property was to emerge through human interaction. Could we perceive and understand it? We might not perceive it simply because we are not looking at it;[12] or perhaps we are looking at it wrongly; or again some of them might simply be imperceptible. These last cases are of no interest: if they have no direct or indirect perceptible effect, they might as well not exist. As for understanding these properties, as we pointed out earlier, a qualitatively new cognitive property might simply be alien to us: it is difficult to imagine such a property, even if we are ready to put that label on properties we would otherwise consider magic, psychic, mystic or something like that (or even worse). Such a property we would perceive but not understand, at least at first. While speaking of swarm intelligence, we also made the following note: ant cognition is not able to use concepts related to the emerging property (short distance, etc.). We have to be open to the possibility that there may be situations where human epistemic limitations prevent the understanding of properties that emerge from our collective behaviour. In such a situation, even when perceived, the essence of the distributed cognitive property will remain out of our intellectual reach.[13]

But the situation is not as bad as it seems. Even if *direct* perception of emergent cognitive properties might be problematic, indirect perception is much more likely. Dark matter cannot be seen, but its effects can be measured. Getting back to our wolf pack, the system-level analysis would be that *Surrounding* is more likely to prevent preys from escaping. At the agent-level (the wolf), the relevant disposition could be described as: "If, while hunting, I stay next to my fellow hunters but not too close, I get more food". Now, some superiorly intelligent wolf that would know of the notion of emergence might suspect that some emergent property is actually having an impact, while not being able to explicitly formulate it. It remains a black box, but assumptions can be made with respect to its inputs and outputs. The same holds for collective cognition: some of the quantitative changes in systemic properties might come not only from aggregation of the same agent-level property, but also from emergent properties' side effects, and these can be perceived.

A last practical thought on distributed and collaborative cognition. In principle, a human with unlimited time and unlimited amount of pen and paper (or a very good memory!) could do everything a computer (or network of computers) can. But as soon as we stop arguing theoretically over the nature of various types of collective cognition, it is obvious that, since all of our capacities, including the cognitive ones, are bounded one way or another, the only option we got to increase cognition is with the help of sidekicks. Hence, the practical importance of studying the interaction and the dynamics of groups of humans cognitively interacting. This is especially true as the heterogeneity of the system's components increases (as in situations where humans and computers cognitively interact), from "simple" situations from the field of Human–Computer Interface (HCI) to that of Computer Supported Collective Work (CSCW), in which the present authors used to collaborate, and all the way to what Licklider called man–computer symbiosis (Licklider 1960).

Notes

1. We use the term here as in first-order fuzzy logic. Both distribution and cognition will be ascribed criteria to determine their membership function, that is, to what extent individuals (here individual properties and systems) belong to the concept's extension.

2. As such, other candidate conditions have been discarded (like communication ability, autonomy, learning or expectations, behaviour based on goals, drives or intentions, qualia, etc.), and the proposed choice is of course open to discussion.

3. By algorithms that set an adequate value to property w of connections.

4. In the case of a MLP, i, j and k shall be adapted in CAT's equation to show that the network is not fully interconnected but that neurons are layered.

5. Of course, the notion of an environment in a neural network is very minimal. When the network is not embodied in a robot or animal, receiving environmental inputs through transducers and manifesting behaviours through effectors, its "environment" is reduced to input and output files, the structure of which rarely possesses typical environmental structure like a topology.

6. And sometimes even with an improvement, as it might prevent overtraining and improve the generalisation capacity of the network.

7. Remember that the agents producing the layout of the footpath have access to cognitive capacities, as they are humans.

8. The swarm approach has other advantages, some related to the multi-agent approach, some related to the choice of simplicity at the agent level. The emergence of global properties from local perceptions and interactions has been shown to be efficient in situations in which the global system is simply unknown or too transient, such as packet routing in networks. The local nature of action and perception in SI systems, coupled with the absence of explicit or complex representation of the world, provides robustness: environment can be changed, even dramatically, agents can be added and subtracted (to some extent: remember this is an emergent property), and chances are that the global properties of the system will not be impaired.

9. Brains, which we haven't discussed here, but which are presumably the ultimate neural network, would of course rate high on both dimensions.

10. As a matter of fact, it can even be seen as a way of constraining the environment by leading the prey and the pack to a defined position, so that other functions can be used without the need for these other functions to be adaptable.

11. Of course, Ockham would not have included *any* properties in his ontology, being a nominalist — nobody's perfect!

12. For example, if we consider steel plates and the Size property, this property is part of every steel plate, but different organisation of the plates — i.e., Wimsatt's CI condition — will change this property at the system level, making it a mildly emerging systemic property. But how would we recognise that some organisation, let's say, being boat-shaped, would provide a qualitative emergent property, Floatability?

13. Of course, if this inability to understand is just a quantitative failure of our cognitive capacities, we can still rely on the aggregation of our efforts (CC) to understand it.

References

Brassac, C. and Pesty, S. 1996. "La pelouse fourmilière; de la coaction à la coopération". In Q. Muller (ed), *4èmes Journées Francophones sur l'Intelligence Artificielle Distribuée et Systèmes Multi-Agents* (JFIADSMA '96). Paris: Hermès, 250–263.

Baron-Cohen, S. 1997. *Mindblindness*. Cambridge, MA: The MIT Press.

Harnad, S. 2005. "Distributed processes, distributed cognizers, and collaborative cognition". *Pragmatics & Cognition* 13(3): 501–514.

Bonabeau, E., Dorigo, M., and Theraulaz, G. 1999. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford: Oxford University Press.

Dorigo, M. and Stützle, T. 2004. *Ant Colony Optimization*. Cambridge, MA: The MIT Press.

Hutchins, E. 1996. *Cognition in the Wild*. Cambridge, MA: The MIT Press.

Licklider, J.C.R. 1960. "Man–computer symbiosis". *IRE Transactions on Human Factors in Electronics* 1: 4–11.

Minsky, M. 1985. *Society of Minds*. New York: Simon and Shuster.

Nonaka, I. and Tekeuchi, H. 1995. *The Knowledge-Creating Company*. Oxford: Oxford University Press.

Wimsatt, W.C. 1986. "Forms of aggregativity". In B. Donagan, A. Perovich Jr., and M.A. Wedin (eds), *Human Nature and Natural Knowledge: Essays Presented to Marjorie Grene on the Occasion of Her Seventy-fifth Birthday*. Dordrecht: Reidel, 259–291.

Authors' addresses

Pierre Poirier
Département de philosophie
Université du Québec à Montréal
C.P. 8888 succ. Centre-ville
Montréal, Québec
Canada, H3C 3P8

poirier.pierre@uqam.ca

Guillaume Chicoisne
Institut des Sciences Cognitives
Université du Québec à Montréal
C.P. 8888 succ. Centre-ville
Montréal, Québec
Canada, H3C 3P8

chicoisne.guillaume@uqam.ca

About the authors

Pierre Poirier is research-professor in the philosophy of cognitive science, Université du Québec à Montréal. Three questions currently interest him. What does mobile robotics teach us about human cognition? How can evolutionary considerations be introduced in cognitive science? What is the best epistemological framework for comprising all systems, from genes to brains to institutions, which properly can be described as possessing/producing knowledge. Living in Montréal, he is happy to have never found it necessary to get a driver's license.

Guillaume Chicoisne is research and planning officer in the Cognitive Science Institute, Montréal. As a proponent of multidisciplinary approaches, his research interests are manifold, but three themes emerge from his past and current projects: virtual communities, multi-agent systems, and man–computer interaction. Nevertheless, his favourite way of working involves mainly pen, paper and a whiteboard.

Distributed cognition

Domains and dimensions

John Sutton
Macquarie University

Synthesizing the domains of investigation highlighted in current research in distributed cognition and related fields, this paper offers an initial taxonomy of the overlapping types of resources which typically contribute to distributed or extended cognitive systems. It then outlines a number of key dimensions on which to analyse both the resulting integrated systems and the components which coalesce into more or less tightly coupled interaction over the course of their formation and renegotiation.

Keywords: distributed cognition, extended mind, memory, social ontology

1. The integration and coordination of domains in distributed cognition

What are the appropriate domains of investigation for research in distributed cognition? In seeing cognitive processes — remembering, reasoning, navigating, planning and so on — as spreading (in certain circumstances) out of the head and into the changing techno-social world, this emerging framework can seem to be too loose and over-inclusive. The objects of a putative science of distributed cognition, complain Adams and Aizawa (2001: 62), would form "an unscientific motley": they note, for example, that Merlin Donald, in his groundbreaking discussion of the history of human use of 'exograms' in external symbol systems, offers rich accounts "of the development of all manner of external representations, including body decorating, grave decorating, sculpture, Stonehenge, hieroglyphics, cuneiform, maps, graphs, and musical scores", and careful analyses of the many "ways in which the processing of exograms differs from the processing of engrams" in the brain (Adams and Aizawa 2001: 58, referring to Donald 1991). And a quick survey of recent work shows the field covering human capacities to manipulate, exploit, or couple

Pragmatics & Cognition 14:2 (2006), 235–247.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

with (for example) other people, scrabble tiles, theatre architecture, cocktail glasses, slide-rules, incised sticks, shells, languages, moral norms, knots, codes, diagrams, fingers, monuments, software devices, rituals, rhythms and rhymes, and roads. This might seem to justify Adams and Aizawa's lament that systems composed of brains coupled with such a diverse range of putatively cognitive tools, computing devices, or memory aids "would seem to form such a motley collection that they will not form the basis for any significant scientific theorizing" (2001: 63).

I am no enemy of motley, but I accept that in response it won't always be enough for distributed cognition enthusiasts to talk of the ecological validity of multidisciplinary immersion in the idiosyncratic and messy reality of cognitive practices. Writing as such an enthusiast, I suggest that — as well as getting on with the work of producing rich case studies and weaving the results into developing sciences of the interface — we must also identify patterns and pick out the most relevant dimensions of comparison across the proliferating studies of "varied, multiplex, interlocking and criss-crossing causal mechanisms" which spread over "a wide variety of mechanistic bases" (Clark, in press). But can we yet offer even the broadest taxonomy of the domains across which cognitive systems are putatively distributed, of the key properties and dimensions on which such interactive systems vary, or even of the existing methods and approaches most effectively used so far in addressing them?

The implicit background to the following blunt and bare sketch of such a taxonomy of domains, dimensions, and methods includes a promiscuous range of variously labelled and subtly differing approaches. David Kirsh (2006: 258) offers us a succinct statement of the overall domain:

> The study of distributed cognition is very substantially the study of the variety and subtlety of coordination. One key question which the theory of distributed cognition endeavors to answer is how the elements and components in a distributed system — people, tools, forms, equipment, maps and less obvious resources — can be coordinated well enough to allow the system to accomplish its tasks.

My remarks apply most centrally to two related frameworks: 'distributed cognition' (DC) as practised by Kirsh and, for example, by Zhang and Norman (1994) and Hutchins (1995); and Andy Clark's 'extended mind' hypothesis (EM) (1997; Clark and Chalmers 1998). But they should apply in broad outline also to core examples of research under the tags 'situated' or 'enactive' or 'embedded' or 'embodied' or 'dynamical' cognition, 'active externalism', or 'vehicle externalism' (Varela, Thompson, and Rosch 1991; van Gelder 1995; Haugeland

1998; Hurley 1998; Rowlands 1999; Dennett 2000; Wilson 2004; Tribble 2005; Wheeler 2005), as well as to related work arising more from independent developments in science studies than from within the cognitive sciences (Latour 1996, 1999; Suchman 1998). In my view the DC/EM movements are in a particularly rich period of cross-tradition and cross-disciplinary interaction, and in these areas it is incumbent on philosophers and others with synthetic and eclectic tendencies to spread and blend relevant theoretical innovations and to catalyse interactions across the difficult gulfs left by specialized training and assumptions.

As Kirsh (2006: 250) notes, because "coordination is the glue of distributed cognition and it occurs at all levels of analysis", particular explanatory projects must aim at system-level approaches to the idiosyncratic interactivity of distributed cognitive systems. But he also acknowledges the occasional utility of artificial analytical separation of even tightly-coupled components, in order to make comparisons and seek generalizations across contexts, and allow the possibility of transferring lessons from any one case study. So we can offer a first brief high-level taxonomy of the overlapping types of resources which typically contribute to and coalesce into such distributed cognitive systems at various timescales.[1]

1.1 External cultural tools, artefacts, and symbol systems

Key DC/EM examples here include the instruments and procedures involved in navigation, or the physical objects and epistemic tools used in processing orders in a café, the tangle of notes and records and processing systems with which an academic paper is written, or the sketchpads without which abstract artists cannot iteratively re-imagine and create an artwork (van Leeuwen, Verstijnen, and Hekkert 1999). Taking a lead from the sociology of material culture, we can call this the study of "the cognitive life of things" (Sutton 2002a, drawing on Appadurai 1986). Of course these resources (at present, in general) do no cognitive work on their own, so it's no argument against a DC approach to emotion, for example, to complain that "the black tie I wear at the funeral isn't doing my grieving for me" (Harris 2004: 729): but then neither do brains tend to do their cognitive work in isolation, because essentially incomplete creatures like us naturally use our neural resources in part to parasitize, lean on, and incorporate those external cultural-technological resources which have become apt for incorporating (Clark 2003). The relevant external resources can include both exograms in various forms of external symbol systems, and different aspects of the nonsymbolic environment (Wilson 2004: 192–197).

1.2 Natural environmental resources

Just as many biological creatures do with *non*-cognitive external resources, organisms can reliably exploit certain kinds of world-mind constancy by functionally integrating environmental structures — such as those exploited in ongoing sensorimotor mastery of couplings between perception and action — to transform their on-board computational tasks and abilities. This creates one typical variety of more-or-less "transient extended cognitive system" (Wilson and Clark, forthcoming).

1.3 Interpersonal and social distribution or scaffolding

Often in some tight complementary fit with technological resources (as in the complex techno-social systems described by Kirsh and Hutchins), other people are more-or-less stably and reliably involved in an individual's cognitive-affective processing. This may occur in some respects among other animals, but the variety and centrality of the interpersonal resources which are integrated into human cognition is one of our most characteristic psychological features. In the case of autobiographical memory, to take a key example, we do on occasion remember alone — in some current neural, emotional, psychological, bodily, and circumstantial context — but the sharing of memories with others is also an ordinary human activity with great psychological and social significance. Sometimes such sharing of memories — like other small-group cognitive activities — is merely additive or aggregative, with each individual bringing fully-formed intact items to the collective arena, communicating them, and taking them away again unaltered: but more often perhaps the social manifestation of memories brings into being new emergent form and content through the transactive nature of collaborative recall (Sutton 2004, 2006; Wilson 2005; Campbell 2006).

1.4 Embodied capacities and skills

Thinking, remembering, feeling, counting and the like may sometimes involve embodied activities in ways which transform the cognitive task and sculpt its process. Embodied interactions with artefacts, or gesturing in characteristic ways in a social situation, or following certain bodily procedures and rituals can (on the DC/EM account) themselves be forms of cognizing, rather than the mere expressions of prior internal cognitive processing (Connerton 1989: Chapter 3; Dreyfus 2002; Anderson 2003; Sheets-Johnstone 2003; Cowart

2004). These embodied cognitive capacities are interwoven in complex ways with our use of the technological, natural, and social resources mentioned above. Although analytically distinguishable, we can also include here the kinds of thinking-in-action apparent in the exercise of certain learned skills in sport, music, and dance (Sutton forthcoming a; Sheets-Johnstone 1999; Sudnow 2001; Stevens, Malloch, McKechnie, and Steven 2003). In these cases occurrent cognitive activity can — in the right circumstances — be distributed across whole patterned sequences of allowable bodily response repertoires, coupling and coalescing dynamically in real time with complex and simultaneous changing physical, technological, and social parameters. For these reasons, expert embodied performance in these domains, and the interactions between kinaesthetic and episodic memory, is a rich and barely-tapped domain of investigation for both ethnographic and cognitive wings of the distributed cognition movements.

1.5 Internalized cognitive artefacts

The four domains discussed so far fit the DC/EM stereotype in that they take the mind beyond the brain, describing psychological states and processes as hybrids, unevenly distributed across social, technological, and biological realms. But two further significant domains of recognizably DC-/EM-styles of enquiry paradoxically enter the traditional stronghold of internalism, the brain, by offering distinctive approaches to the internal wing of extended cognitive systems. On the one hand, we may get a new grip on the nature and unique contribution of the relevant neural resources: analysis of the complex wholes made up when embodied brains couple with 'cognition-amplifiers' like objects and other people "may itself contribute important insights concerning the contributions and functioning of the biological brain itself" (Wilson and Clark, forthcoming). As our independent investigations of the dynamic neural bases of memory, reasoning, action, and so on continue, they will be enriched by direct evidence of the alterations in and constraints on processing imposed by interaction with particular kinds of external symbols or social structures (Clark 2005a). This is just to underline the post-connectionist heart of the DC/EM picture: it's just because items of information are *not* in general held in stable and discrete form in the brain that we so pervasively lean on the bits of the world which we have individually and collectively made smart. Secondly, an important strand of DC/EM work — as ever, with complex historical roots in Vygotsky and elsewhere — addresses internalized versions of external or cultural resources. We use a wide range of stratagems to bootstrap, manage, transform, and discipline our minds, and these techniques can coopt internal surrogates as worldly exograms. Both linguistic items — words, labels, phrases — and other symbols can play key cognitive roles independent of any communicative function, in freezing thoughts or condensing complex affects, or in becoming (for example) "a new [internal] target for selective attention and a new fulcrum for the control of action" (Clark 2006: 294). One rich real-world example of such complex acquired internal cognitive architectures was the baroque Renaissance art of memory, through which monks and scholars learned rhetoric and meditation and trained their wandering minds in the craft of thought (Carruthers 1998). The virtual random access systems and internal prostheses which these adepts built into their theatres of memory were extended as well as cognitive, even though they didn't happen to be outside the skull and skin (Sutton In press). Such forms of "virtuoso artificial self-manipulation" can be dramatically culturally specific and yet still a natural structuring supplement to the cognitively-incomplete biological brain, vital cognitive-affective resources with which "to drive, sculpt, and discipline the internal representational regime" (Clark 2005b: 264). Among other things, internalized cognitive artefacts — linguistic, imagistic, or other — can add an extra layer of insulating variation against the dynamics of world and mind. The context-sensitivity and openness to external influence of remembering, feeling, and thinking is itself context-sensitive: under certain conditions, we may temporarily approximate more detached cognitive agents, dynamically buffered from the world to recalibrate or renegotiate our memories, plans, or emotions.

2. Dimensions of distribution and integration

This selective tour of some domains of distributed cognition — domains which in everyday contexts are often tangled together — now needs to be supplemented with at least an initial account of what we're looking for in comparing different cases and kinds of the resulting extended cognitive systems. So far we have catalogued some differences in the nature of the extended resources which the integrating mind coopts in distributed cognitive activity. There are also a range of possible dimensions of difference in the nature of the resulting extended cognitive system. What are some typical dimensions of analysis?

As a remembering or planning or reasoning or feeling or navigating creature, I may gain quite different capacities and idiosyncrasies (and patterns of breakdown) when I hook up with specific other people or objects or environments, or when I train myself in and rely on specific learned bodily or cognitive techniques. In many cases I just couldn't perform in the same way if I could lean on the unsupported capacities of my naked brain; and even when I *could* perform the relevant tasks 'on my own' (accessing only the internalized surrogates of these external resources), the way in which this potential performance is actualized will often shift as my openness to influences of various kinds alters.

Kirsh (2006) asks for better formalisms to understand the "principles of coordination" of the different components in a distributed system. In a related development of systems thinking, Haugeland (1998) questioned the idea of determinate interfaces between apparently separable components when a system is interconnected at sufficiently high bandwidths. And Poirier and Chicoisne (2006) focus on a set of related dimensions by which to measure the extent of emergence — understood as the failure of aggregativity — in any system. On their view, more truly distributed systems have components that are more tightly integrated or coupled, with a specific organization of bushy interactivity which makes a significant contribution to overall functioning. These dimensions of emergence are important signs of the "continuous reciprocal causation" between (for example) brain, body, and world, which is sometimes identified as the mark of dense reciprocal cognitive coupling (Clark 1997: 163–166).

Alongside these marks of distribution-as-emergence, we can also follow Wilson and Clark (forthcoming) in stressing further diachronic dimensions of variation. They point to the importance of "the durability and reliability of the extended cognitive system that results from the functional integration" of extended resources. Some distributed systems are one-offs, establishing transient and easily-dissoluble relations which are still more-or-less controllable and reliable (and which may still exhibit high degrees of emergence). Perhaps more common in human cognition is the soft assembly of transient but regularly repeatable integrated wholes involving both internal and external forms of representation: our neural resources may come, in the course of development, to be "expressly tailored to accommodate and exploit the additional representational and computational potentials introduced by, for example, the compass, the pen and paper, or the word-processing package". Wilson and Clark then characterize the more dramatic notion of the 'extended mind' as simply any cognitive extension which scores highly on these dimensions of durability and reliability, in which whatever new capacities emerge in the coupled system are "sufficiently robust and enduring as to contribute to the persisting cognitive profile" of a specific individual.

These are all extremely useful suggestions which can be mapped onto existing case studies in distributed cognition, and which all take distribution as a multidimensional matter of degree. In conjunction with the diversity of types of relevant external resource surveyed in Section 1, these suggestions already demonstrate that — contrary to some critics (Butler 1998: 211–212; Adams and Aizawa 2001: 55–59) — the DC/ EM frameworks do not necessarily treat internal and external resources as on a par or identical. Rather, the various components in a distributed cognitive system usually complement but do not replicate each others' characteristics (Clark 1997: 220; in press). The central line of thought is thus not based on the parity between neural and environmental resources, but on their (more-or-less tight, more-or-less durable) complementarity (Menary 2006; Sutton, in press).

Once complementarity is installed as the core motivation, the DC/EM framework can be understood as investigating a range of relationships between engrams and exograms, or between agents and artefacts, relations which may be asymmetric, tangled, and dynamically reconfigured or renegotiated. This allows us to investigate particular characteristics of both sets of resources. Different non-biological external resources (technologies, media, other people, places, and so on) vary on a whole range of dimensions including not only their durability but also (for example) the medium-dependence or translatability of the information they carry, and its transmissibility across instantiations, their capacity as symbol systems, the constraints on the ways in which information can be retrieved from them, the context-dependence of their use and so on (compare Donald 1991: 315–316).

Mainstream DC/EM work has tended to treat all systems of exograms, in particular, as similarly 'classical' in format — passive, stable, and medium-independent, in contrast to active, reconstructive biological memory. But of course (as art historians and archaeologists remind us) not all external memory fields or systems are meant to be so permanent, and not all which are intended to endure actually do so (Kwint 1999; DeMarrais, Gosden, and Renfrew 2004). Future technologies may familiarize us better with external resources at least as dynamic and context-dependent as other people in our socio-cognitive world already are. But equally, historical cognitive science (Sutton, in press) can confirm the contingent and multiple nature of possible extended cognitive physiologies: if we are cyborgs *by nature*, we have never been "bound and restricted by the biological skin-bag, … the ancient fortress of skin and skull" (Clark 2003: 4–5; compare Latour 1993). So history needs to take its place alongside ethnography and developmental psychology as a key testing-ground for the whole framework (Tribble 2005; Sutton, forthcoming b). This attention to detail about the properties of exograms and other external resources stops our investigations of inner-outer relations relying on either an assumption of parity or a sharp dichotomy between fluid biology and stable culture: as Hutchins (1995:

312) argued, "it is not that some content is copied from the outside world into some internal storage medium ... what used to look like internalization now appears as a gradual propagation of organized functional properties across a set of malleable media". Inner and outer realms are not pre-divided into natural and artificial, each with its own inevitable proprietary characteristics: rather, such boundaries are hard-won and fragile developmental and cultural achievements, always open to renegotiation.

And yet, finally, we also want to understand whatever diachronic stability and continuity particular embodied agents exhibit. As I go round in a complex world of varied external resources, I can move between and enter into distinctive relations with quite different artefacts and other agents. Each of us can decouple from and recouple with external resources of various kinds on a regular and continuing interactive basis. If parity considerations exclusively drove DC/EM thinking, so that features of inner and outer resources tended to be analytically collapsed into each other, then these facts would be mysterious (Butler 1998: 208–210; Grush 2003: 79–81). But highlighting the complementarity between distinct inner and outer resources of course allows for such tracking of agents over time and across their transient material and social engagements. So the last set of dimensions we need to keep open for analysis are features of the individual differences in the ways people approach various cognitive tasks, some characteristically *without* significant use of external resources. Integration with mainstream personality and social psychology here is long overdue. Even in tasks which can involve extended looping and coupling cognitive processes, we are all familiar with folk who aren't content or able to leave the information out in the world, or to use the world as its own memory. In developing Brooks's anti-representationist arguments, John Haugeland claimed that "it would be silly, for most purposes, to try to *keep track of* what shelf everything in the refrigerator is currently on; if and when you want something, just *look*" (1998: 219). But, I suggest, we all know people who *do* typically upload such information into their on-board biological memories: such individual differences in the amount and style of reliance on external resources are often glaring in the ways people plan and engage in complex activities, such as writing an academic paper, shopping for a party, or chairing a department meeting. Do I memorise the train timetable in advance, or do I just turn up to the station and see? These matters of personality and psychological style in the distribution of cognitive resources can also have considerable normative and moral weight.

There is no space here to detail just how these kinds of dimensions are indeed already under investigation, within broadly DC/EM frameworks, in a number of disciplines and traditions with their own independent histories and methods. In developmental psychology, to finish with just one further example, the robust empirical tradition of social-interactionist research on early autobiographical memory already attends closely to the transmission of social ontologies and to both cultural and individual differences, within an overall dynamical developmental systems framework (Griffiths and Stotz 2000; Reese 2002; Sutton 2002b; Nelson and Fivush 2004). I've indicated above that, in my view, the DC/EM framework can already also draw on and in turn illuminate independent work in (for example) social ontology within philosophy, cognitive archaeology and studies of material culture, media theory, a range of historical disciplines, personality psychology, studies in the social-cognitive psychology of (for example) collaborative recall and transactive group cognition, and theories of embodied expertise and skill. Interdisciplinarity in the cognitive sciences, of course, has its pitfalls. But amidst the vast apparatus and social-institutional weight of the Kuhnian 'normal science' which is deeply embedded in more mainstream modern cognitive and neurocognitive sciences, perhaps some attention to the messy nature of our shared social and cognitive world might occasionally excuse some proliferating cross-disciplinary indulgence of the current thrills and promises within these cross-disciplinary approaches to distributed cognition.

Note

1. My scheme here owes much to that offered by Wilson and Clark (forthcoming), though mine is more elaborate and inclusive, and less fully detailed. Note that this is not a taxonomy of the dimensions which might count towards *making* a system either distributed or cognitive, along the lines of Poirier and Chicoisne's (2006) set of conditions. Here I'm just taking for granted the existence of some such systems, and in particular not addressing any concerns about relations between phenomenal consciousness and (distributed) cognition (but see for example Rowlands 2003: Chapter 10). I am entirely in sympathy with Poirier and Chicoisne's important claim that the concept of distribution is continuous or fuzzy rather than all-or-nothing, and the dimensions of variation which I identify in Section 2 are a start at a related and compatible way of identifying some further components of that fuzzy concept.

References

Adams, F. and Aizawa, K. 2001. "The bounds of cognition". *Philosophical Psychology* 14: 43–64.

Anderson, M.L. 2003. "Embodied cognition: a field guide". *Artificial Intelligence* 149: 91–130.

Appadurai, A. 1986. *The Social Life of Things: Commodities in Cultural Perspective.* Cambridge: Cambridge University Press.

Butler, K. 1998. *Internal Affairs: Making Room for Psychosemantic Internalism.* Dordrecht: Kluwer.

Campbell, S. 2006. "Our faithfulness to the past: reconstructing memory value". *Philosophical Psychology* 19: 361–380.

Carruthers, M. 1998. *The Craft of Thought.* Cambridge: Cambridge University Press.

Clark, A. 1997. *Being There: Putting Brain, Body, and World Together Again.* Cambridge, MA: The MIT Press.

Clark, A, 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence.* Oxford: Oxford University Press.

Clark, A. 2005a. "Beyond the flesh: some lessons from a mole cricket". *Artificial Life* 11: 233–244.

Clark, A. 2005b. "Word, niche and super-niche: How language makes minds matter more". *Theoria* 20: 255–268.

Clark, A. 2006. "Material symbols". *Philosophical Psychology* 19: 291–307.

Clark, A. In press. "Memento's revenge: The extended mind extended". In R. Menary (ed), *The Extended Mind.* Aldershot: Ashgate.

Clark, A. and Chalmers, D. 1998. "The extended mind". *Analysis* 58: 7–19.

Connerton, P. 1989. *How Societies Remember.* Cambridge: Cambridge University Press.

Cowart, M. 2004. "Embodied cognition". *The Internet Encyclopedia of Philosophy.* http://www.iep.utm.edu/e/embodcog.htm

DeMarrais, E., Gosden, C., and Renfrew, C. 2004. *Rethinking Materiality: The Engagement of Mind with the Material World.* Cambridge: McDonald Institute for Archaeological Research.

Dennett, D.C. 2000. "Making tools for thinking". In D. Sperber (ed), *Metarepresentations: A Multidisciplinary Perspective.* Oxford: Oxford University Press, 17–29.

Donald, M. 1991. *Origins of the Modern Mind.* Cambridge, MA: Harvard University Press.

Dreyfus. H.L. 2002. "Intelligence without representation: The relevance of phenomenology to scientific explanation". *Phenomenology and the Cognitive Sciences* 1: 367–383.

van Gelder, T. 1995. "What could cognition be, if not computation?" *Journal of Philosophy* 92: 345–381.

Griffiths, P.E. and Stotz, K. 2000. "How the mind grows: A developmental perspective on the biology of cognition". *Synthese* 122: 29–51.

Grush, R. 2003. "In defense of some "Cartesian" assumptions concerning the brain and its operation". *Biology and Philosophy* 18: 53–93.

Harris, R. 2004. "Integrationism, language, mind and world". *Language Sciences* 26: 727–739.

Haugeland, J. 1998. "Mind embodied and embedded". In *Having Thought: Essays in the Metaphysics of Mind.* Cambridge, MA: Harvard University Press, 207–237.

Hurley, S.E. 1998. *Consciousness in Action.* Cambridge, MA: Harvard University Press.

Hutchins, E. 1995. *Cognition in the Wild.* Cambridge, MA: The MIT Press.

Kirsh, D. 2006. "Distributed cognition: A methodological note". *Pragmatics & Cognition* 14(2): 249–262.

Kwint, M. 1999. "Introduction: the physical past". In M. Kwint, C. Breward, and J. Aynsley (eds), *Material Memories.* Oxford: Berg, 1–16.

Latour, B. 1993. *We Have Never Been Modern.* Cambridge, MA: Harvard University Press.

Latour, B. 1996. "Cogito ergo sumus! Or, psychology swept inside out by the fresh air of the upper deck: review of Hutchins 1995". *Mind, Culture, and Activity* 3: 54–63.

Latour, B. 1999. "A collective of humans and nonhumans". In *Pandora's Hope: essays on the reality of science studies.* Cambridge, MA: Harvard University Press, 174–215.

van Leeuwen, C., Verstijnen, I., and Hekkert, P. 1999. "Common unconscious dynamics underlie uncommon conscious effects: a case study in the interaction of perception and creation". In J. Jordan (ed), *Modeling Consciousness Across the Disciplines.* Lanhan, MD: University Press of America, 179–218.

Menary, R. 2006. "Attacking the bounds of cognition". *Philosophical Psychology* 19: 329–344.

Nelson, K. and Fivush, R. 2004. "The emergence of autobiographical memory: A social cultural developmental theory". *Psychological Review* 111: 486–511.

Poirier, P. and Chicoisne, G. 2006. "A framework for thinking about distributed cognition". *Pragmatics & Cognition* 14(2): 215–234.

Reese, E. 2002. "A model of the origins of autobiographical memory". In J.W. Fagen and H. Hayne (eds), *Progress in Infancy Research,* volume 2. Hillsdale, NJ: Erlbaum, 215–260.

Rowlands, M. 1999. *The Body in Mind: Understanding Cognitive Processes.* Cambridge: Cambridge University Press.

Rowlands, M. 2003. *Externalism.* Chesham: Acumen.

Stevens, C. Malloch, S., McKechnie, S., and Steven, N. 2003. "Choreographic cognition: The time-course and phenomenology of creating a dance". *Pragmatics and Cognition* 11: 299–329.

Suchman, L. 1998. "Human/ Machine Reconsidered". *Cognitive Studies* 5: 5–13.

Sudnow, D. 2001. *Ways of the Hand: A Rewritten Account.* Cambridge, MA: The MIT Press.

Sheets-Johnstone, M. 1999. *The Primacy of Movement.* Amsterdam: John Benjamins.

Sheets-Johnstone, M. 2003. "Kinesthetic memory". *Theoria et Historia Scientiarum* 7: 69–92.

Sutton, J. 2002a. "Porous memory and the cognitive life of things". In D. Tofts, A. Jonson, and A. Cavallaro (eds), *Prefiguring Cyberculture: An Intellectual History.* Cambridge, MA & Sidney: The MIT Press & Power Publications, 130–141.

Sutton, J. 2002b. "Cognitive conceptions of language and the development of autobiographical memory". *Language and Communication*: 375–390.

Sutton, J. 2004. "Representation, reduction, and interdisciplinarity in the sciences of memory". In H. Clapin, P. Staines, and P. Slezak (eds), *Representation in Mind: New Approaches to Mental Representation.* Amsterdam: Elsevier, 187–216.

Sutton, J. 2006. "Integrating the philosophy and psychology of memory: Two case studies". In M. de Caro, F. Ferretti, and M. Marraffa (eds), *Cartographies of the Mind: The Interplay between Philosophy and Psychology.* Dordrecht: Kluwer, 69–79.

Sutton, J. In press. "Exograms and interdisciplinarity: history, the extended mind, and the civilizing process". In R. Menary (ed), *The Extended Mind.* Aldershot: Ashgate.

Sutton, J. Forthcoming a. "Batting, habit, and memory: The embodied mind and the nature of skill". *Sport in Society.*

Sutton, J. Forthcoming b. "Spongy brains and material memories". In M. Floyd-Wilson and G. Sullivan (eds), *Embodiment and Environment in Early Modern Europe.* London: Palgrave Macmillan.

Tribble, E. 2005. "Distributing cognition in the Globe". *Shakespeare Quarterly* 56: 135–155.

Varela, F., Thompson, E., and Rosch, E. 1991. *The Embodied Mind.* Cambridge, MA: The MIT Press.

Wheeler, M. 2005. *Reconstructing the Cognitive World: The Next Step.* Cambridge, MA: The MIT Press.

Wilson, R.A. 2004. *Boundaries of the Mind.* Cambridge: Cambridge University Press.

Wilson, R.A. 2005. "Collective memory, group minds, and the extended mind thesis". *Cognitive Processing* 6: 227–236.

Wilson, R.A. and Clark, A. Forthcoming. "How to situate cognition: letting nature take its course". In M. Aydede and P. Robbins (eds), *The Cambridge Handbook of Situated Cognition.* Cambridge: Cambridge University Press.

Zhang, J. & Norman, D.A. 1994. "Representations in distributed cognitive tasks". *Cognitive Science* 18: 87–122.

Author's address

John Sutton
Department of Philosophy
Macquarie University
Sydney
NSW 2109
Australia

jsutton@scmp.mq.edu.au
http://www.phil.mq.edu.au/staff/jsutton/

About the author

John Sutton is currently head of the Philosophy Department at Macquarie University in Sydney. He is interested in the interdisciplinary study of memory, and in testing the methodological implications of distributed cognition and the 'extended mind' hypothesis by seeing if these frameworks have anything to offer memory researchers in a range of other disciplines. His other interests include Descartes, animal spirits, habit and skill, dreaming, the history and philosophy of science, and the foundations of cognitive science.

Distributed cognition

A methodological note*

David Kirsh
University of California San Diego

Humans are closely coupled with their environments. They rely on being 'embedded' to help coordinate the use of their internal cognitive resources with external tools and resources. Consequently, everyday cognition, even cognition in the absence of others, may be viewed as partially distributed. As cognitive scientists our job is to discover and explain the principles governing this distribution: principles of coordination, externalization, and interaction. As designers our job is to use these principles, especially if they can be converted to metrics, in order to invent and evaluate candidate designs. After discussing a few principles of interaction and embedding I discuss the usefulness of a range of metrics derived from economics, computational complexity, and psychology.

Keywords: epistemic actions, coordination, interaction, design theory, ethnography

1. Introduction

People interact with artifacts, technologies, surfaces and other people in ways that overwhelm our current formalisms. Dynamical systems theory, game theory, economic theory, the formal theory of distributed computation, all fall short of providing satisfying explanations. Formalisms based on abstractions such as task environments fare no better. They are known to be flawed in making unreasonable assumptions about how rational people are, about how the environment of action can be characterized as a fixed set of choice points with fixed option sets rather than a place where people dynamically engage the world interactively. Look at any realistic environment — offices, cafés, computer 'desktops' — all these defy reduction to layerings or intersections of task environments. How are we to proceed in our search for a scientific study?

Pragmatics & Cognition 14:2 (2006), **249–262**.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

Despite the importance of taking a system view, my own approach is to move from the individual to the group. If we can understand how individual people engage their environments, how they appropriate artifacts, how they rely on material aspects of their activity space to help them stay in control, to manage thought, perception and choice, then perhaps we can begin to put these individuals together into larger socio-technical systems. It is not easy to predict how individual behavior changes once people are put together. Coordinating forces, at the system level, help constrain individuals; but in the end people still act locally. Distributed cognition remains a challenge precisely because it is so hard to balance the reality of local choice with system constraint. We need better theories of how people are embedded in the world as well as better theories of how the world and the larger systems we are part of coordinate action.

The following six assumptions have guided my own research in looking for principles of interaction that might fit this bottom-up top-down model.

1. We act locally and are closely coupled to our local environments.
2. We externalize thought and intention to harness external sources of cognitive power.
3. Economic metrics have a place in evaluating distributed systems, but they must be complemented with studies of computational complexity, descriptive complexity and new metrics yet to be defined.
4. The best metrics apply at many levels of analysis, from the system level where our concern is with the goodness of a system's design to the level of individual artifacts, where our concern is with the goodness of the design of the artifacts individuals interact with.
5. Coordination is the glue of distributed cognition and it occurs at all levels of analysis.
6. History matters.

My objective in this note is to introduce each idea in a personal way, as an opportunity to open a dialogue with designers interested in shaping interactivity, which is one of the topics of this special issue.

2. Closely coupled

Let us say that two entities are closely coupled if they reciprocally interact: changes in one cause changes in the other, and the process goes back and forth in such a way that we cannot explain the state trajectory of the one without looking at the state trajectory of the other. When a person writes on paper, the two form a reciprocal system. The person causes paper changes, paper changes partially cause person changes. This reciprocal interaction allows the person to find expressions, to represent and explore ideas using the persistent state of the paper that would otherwise be impossible. There is a dynamic between the two.

One of my first appreciations of the practical value of seeing cognition as distributed, the outcome of a close coupling with external processes, came from studying Tetris (see Figure 1). In work on Tetris, Paul Maglio and I found that players had to be closely coupled to the game in both an epistemic and pragmatic way to play well (Kirsh and Maglio 1994).

The pragmatic side of coupling is uncontentious. In video games, where what happens next depends substantially on what a player does a moment before, interaction is intense, high frequency and the two, player and game, are closely coupled. They form a distributed system whose trajectory cannot be understood without explaining the role each plays in driving the next state.

In Tetris, however, we found that it was not possible to *understand* the actions players took if we looked just at their pragmatic goals in playing. They seemed to be coupled to the system epistemically too, often performing actions not for pragmatic advantage but rather for epistemic reasons. For instance, given a choice between performing a mental action of rotating a mental Tetris piece or a physical action of rotating a physical Tetris piece they often chose the physical action. The net epistemic effect of knowing what a piece would look

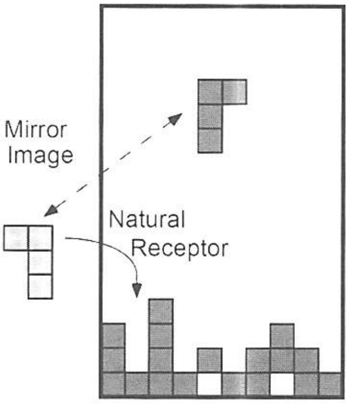


Figure 1. In Tetris the goal of play is to relentlessly fill gaps on the bottom layers so as to complete rows. The game ends when the board clogs up and no more pieces can enter. Because there are mirror pieces, and the choice of where to place a piece is strategic, Tetris players need many hours of practice to become expert. We found that players often rotate pieces to speed up their identification process, especially among pieces with mirror counterparts.

like rotated 90° or 180° would be the same. But physical rotation was faster than mental rotation so the same epistemic state could be reached faster if done physically. Physical rotation would have been pointless if it were not coupled so closely in time with thinking.

Our analysis let us say that these 'extra' moves paid enough in *epistemic* benefits to cover their pragmatic costs. In Tetris we were able to use speed accuracy of judgments and the impact of those judgments on game score to develop a metric for evaluation. Thus benefits, measured in terms of increased speed of piece recognition, improved choice of placement location, and improved control and visual judgment, more than outweighed costs, measured as the number of steps off the most direct path and the time to recover.

To mark this distinction in the role of actions we called intentional movements taken to bring a subject physically closer to its external goals *pragmatic actions* and those intended to simplify computation, reduce error or increase precision, *epistemic actions*. The distinction was not hard and fast. Certain pragmatic actions could also serve epistemic ends. And epistemic actions could be seen as pragmatic with respect to advancing epistemic ends when these were the external goal. But the point was that it is not always obvious why someone does something. We may think they are very goal directed in a simple sense. But without a better theory of exactly what is going on in their head and their world we might easily approach a task and mischaracterize the actions being taken. Actions might be either pragmatic or epistemic. Therefore the metric to evaluate action strategies, artifacts and technologies, may have to have pragmatic and epistemic components and be completely non-apparent.

Epistemic actions are everywhere. Some are connected with uncovering information. We move our hands close to something hot before grasping it. In woodworking the adage is measure twice cut once. Some epistemic actions compensate for sensory limitations. A trick most of us learned as kids was to squint our eyes or make a little hole with our fingers to look through in order to see distant items more clearly. Others are interactive strategies for externalizing. Say something quietly out loud to make sure it is coming out right — this serves to confirm. Some serve as reminders. Thoughtfully place one's keys in front of the door or in one's shoes to save relying on prospective memory alone. There are many other forms of epistemic actions. All have personal payoffs and depend on interaction with the environment.

The way to discover epistemic actions, as with most close interactions actions, is to record activity by video or computer and then ethnographically analyze it. We did this with Tetris and we also data mined our records of agent-environment activity. But in contexts where there is no predetermined metric

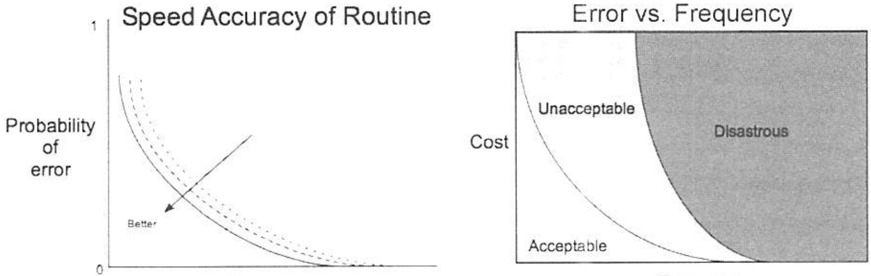
— in activities other than playing video games, for instance — it is hard to measure cognitive and physical costs and benefits.

3. Cost structure

My own efforts to develop metrics for evaluating the costs and benefits of actions, resources, and environments draw on ideas from complexity theory and economics. Consider economic measures first.

In recent ethnographic work at a local café I will call Jolt's, I have been studying the subsystem for taking customer orders and communicating them to the baristas who make the drinks. To develop metrics I thought it would be helpful to contrast the method at Jolt's with the method at another café which I will call Buck's. Buck's still relies on a paper cup that is supposed to be written on; Jolt's uses an IBM ordering system, with an overhead monitor displaying the queue of orders. Both are excellent mechanisms for coordinating a complex system of parts and processes, but each has strengths and weaknesses. The goal of the study was to develop objective ways of measuring some of these strengths and weaknesses.

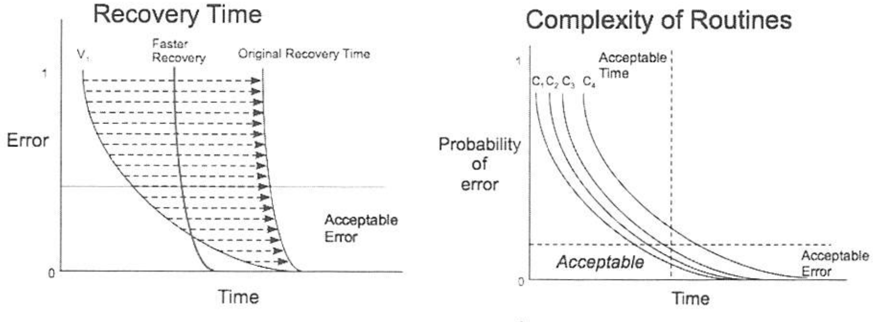
I considered six metrics: speed accuracy, error type vs. frequency, error recovery rate, variance, learnability, and drink complexity. See Figures 2, 3, and 4.



a.

b.

Figure 2 a. A speed accuracy graph shows the average time it takes skilled users to place an order vs. the probability that their entry is wrong. As users get better at their job we assume that they reach a maximum level of performance that is close to some ideal level given the technology, artifacts and procedures they are operating with. A better techno-social system ought to translate into better speed accuracy curves.
b. Because errors come in different flavors with different consequences, they need to be categorized, assessed for frequency, and their probable impact evaluated.

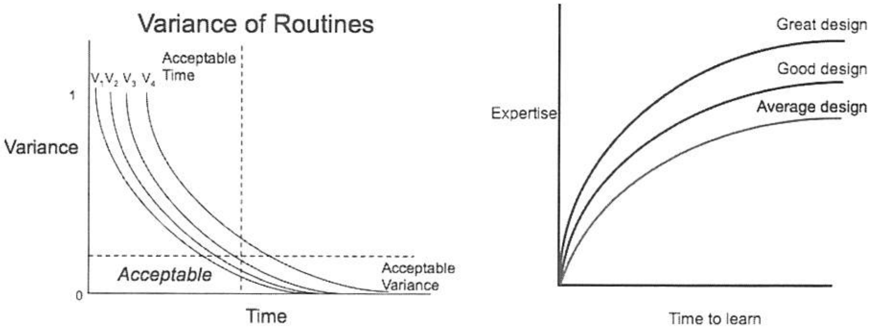


a. b.

Figure 3 a. A natural way to measure the value of the resilience of a system is to observe the time it takes the system to recover from an error, an interruption, or an error in process, and use that measure to compare techno-social systems.
b. At modern cafés drink complexity has risen dramatically. An excellent technology such as the Buck's cup need not increase production speed if it permits staff to produce drinks of greater complexity in acceptable time and quality.

Anyone interested in the evolution and spread of routines in cafés or other organizations would do well to compare routines using graphs and measures as in Figures 2 and 3. But there are a few things to appreciate about these sorts of economic analyses.

First, each graph measures a different parameter. This raises the question: how do we decide if it is good to introduce an artifact that improves speed accuracy but is worse under some other measures such as being harder to learn?



a. b.

Figure 4 a. A basic attribute of a good design is that it reduces the variance of output. The holy grail of quality control is uniformity.
b. Both Jolt's and Buck's have relatively high turnover in staff. Other things being equal, one system is better than another if new staff can more quickly learn the ordering and espresso making system, reaching the same performance level sooner.

If there is no independent measure of value for comparing the relative goodness of different parameters, then we must rely on a Pareto style assessment where all that can be said is: Env_1 is better than Env_2 only if Env_1 is at least as good as Env_2 in every parameter and better in at least one. This hardly addresses the reality of tradeoffs in design.

Second, these sort of quantitative abstractions don't explain *why* Env_1 is better. We have no mechanism. Thus in saying that the Buck's cup yields a better speed accuracy and error recovery rate than the Jolt's monitor system, we have *not* said what about the cup and monitor system make it so. For that we need a microanalysis of the type distributed cognition has become well known for (see, for example, Hutchins 1995). But ethnographic studies tend not to lead to generalizations about mechanisms — which seems necessary for a science of design.

Third, the strength or weakness of a technology will not be apparent unless we also include a careful account of what Vicente (2004) has called its 'human technology'. An example of human technology is the protocol required when taking an order. In Jolt's, as in Buck's, cashiers are required to repeat the drink specification to the customer. This recruits the customer's help in verifying the entry and has as a side effect that an alert barista, if nearby, can anticipate the drink. This formal practice reduces error. But practice is a complex phenomenon in itself. Metrics that are used to measure how good a technology is leave unexplained the interdependence of technology and practice. Performance is assumed to vary with skill (or learnability) but nothing is said about the nature of this skill.

To quantify practices I have tried, at times, to approach them as computational systems — routines really — that involve procedures, memory, control of attention, skill and reasoning. If a practice can be interpreted computationally then we can measure its complexity. How fast do memory requirements, the number of steps involved, and the number of items to monitor, rise as the size of the task is increased? How do cashiers cope with very large orders, or with two people ordering at once, or with one person giving two orders? What is the complexity of the momentary computational problems they face? Can they reduce these problems by doing things like making annotations as they work, by laying out physical objects to encode hints or cues? Do they have a way of talking to clients that lets them enter an order in a step by step manner so that they don't have to remember as much at each moment, or so that they are less likely to make an error while dual tasking?

I undertook a variety of these small computational analyses (Kirsh 1995, 1996). See Figure 5 for another example of how we can measure the effect of

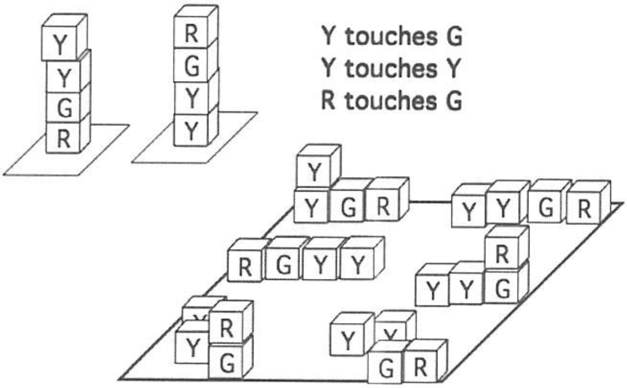


Figure 5. Just changing the size of the mat on which a problem has to be solved can change its complexity. In this case the blocks must be arranged so that the following constraint is met: Y touches G, Y touches Y, R touches G. Solving this problem on a small mat restricts the solution to the two towers shown. Working on a large mat however increases the solution space because now there are two dimensions along which the constraints can be met.

simple environmental manipulations on the complexity of a task. Tasks can be inventively changed to change their complexity profile. The question is: do participants reshape their tasks and environments this way? At this point ethnography becomes indispensable. Of the many things that participants might do to save cognitive effort, which are the ones they actually do and why? For instance, if a computational analysis suggests that a given practice should be difficult to master, and if ethnography shows that it is nonetheless prevalent, then we now have a specific question to ask: what is it about the way people use their resources that makes this problem easier than we assumed? Or is it done for historical reasons, because it is the way people always have?

4. Cognitive efficiency of design

My argument so far has been that, although economic metrics are helpful and necessary to analyze techno-social systems, it is necessary to observe the way people do things in ever more detail and tie these actions to cognition and computational analyses to get at the mechanisms of distributed cognition. The same type of analyses can be used to begin to measure and explain the effectiveness and efficiency of an artifact's design.

To develop deeper intuitions about the effectiveness and cognitive efficiency of designs — in particular visual designs — examine the following two figures. Why is Figure 7 better than Figure 6?

To answer this sort of question we need to develop cognitively motivated principles of design. What drives the way people interact with artifacts? In Figures 6 and 7 the point of the artifact is to structure and present forced choices. Any design that makes the structure and set of choices easier to appreciate, more visible, is a better design. The visibility principle at play here is 'What goes together semantically should go together visually'.

In Figure 6 the space between choices, labels and categories is so *ad hoc* that it is not easy to scan the figure and quickly say without looking at the labels what group of radio buttons one is choosing between. There is no clustering, no use of Gestalt principles of grouping to reflect in the layout the options which belong together. It is true that alignment is used to indicate a connection between category (e.g., Family) and options (e.g., Courier). But if we take as measures of efficiency the time it takes to scan this image and formulate a simple plan, the time it takes to determine if one has completed the entire form, and the time it takes to decide how much is left to be done, all will show that Figure 6 is less efficient than Figure 7.

If we were to give metrics for comparing two designs they would look much like our economic measures. For example, since subjects must decide if they have completed their task we can create a simple decision task and experimentally compare the speed accuracy curves for Figures 6 vs. 7. We can measure what types of errors are made and how frequently. Similarly we can

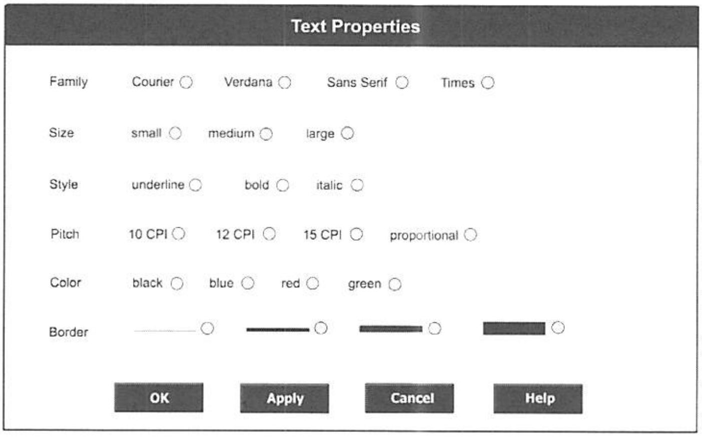


Figure 6. Here is a dialogue box for setting the text properties of a style in a word processor like MS Word in the 80's and the days of Microsoft DOS.

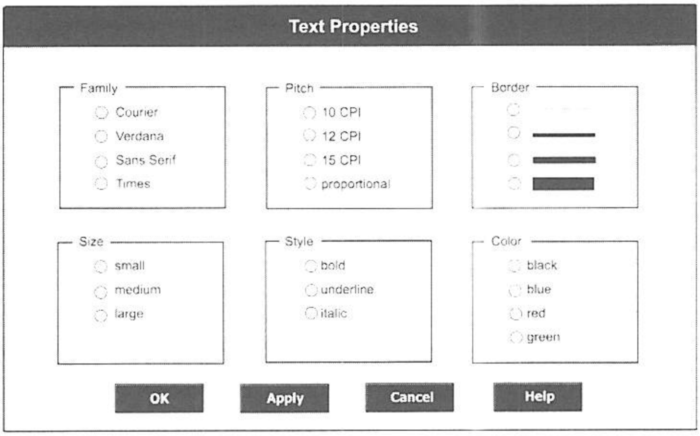

Text Properties
Family
○ Courier
○ Verdana
○ Sans Serif
○ Times
Pitch
○ 10 CPI
○ 12 CPI
○ 15 CPI
○ proportional
Border
Size
○ small
○ medium
○ large
Style
○ bold
○ underline
○ italic
Color
○ black
○ blue
○ red
○ green
OK
Apply
Cancel
Help

Figure 7. The same labels have now been distributed more effectively, using borders and positioning to simplify the perception of groups. The speed accuracy measures of performance on this layout trump those of Figure 6.

explore cognitive costs through computational analysis: suppose the problem were increased in size (more parameters to set). Are there more efficient algorithms available for one of the layouts than the other?

5. Coordination

The study of distributed cognition is very substantially the study of the variety and subtlety of coordination. One key question which the theory of distributed cognition endeavors to answer is how the elements and components in a distributed system — people, tools, forms, equipment, maps and less obvious resources — can be coordinated well enough to allow the system to accomplish its tasks. Even coordinating mechanisms as simple as clocks or paper clips can make the difference between a successful system and an unsuccessful one. Clearly we would like methods and measures for systematically exploring coordination.

Part of my own pleasure in studying humans comes from discovering the remarkably diverse ways we coordinate activity. So far I have briefly mentioned how individuals coordinate different parts of cognition by exploiting the environment. People manipulate local conditions to stay in control, to perform faster and more effectively. They annotate to cue response and reduce prospective memory, they line up items to make them easier to scan, to notice outliers and so on.

These same sorts of principles are exploited by good designers when they make artifacts that make our life easier. But all these examples of coordination are local; they affect local choice. In distributed systems the success of the whole depends equally on all these acts of local choice adding up, working together to move the system closer toward system goals. In a café it's about the beverages not the baristas. The system works well if clients get a satisfying drink, the one they ordered, quickly, at the right price, and well made. For that to happen, everyone's roles must fit.

Decisions about the roles people will play in a system, like the decisions about the artifacts, physical layout, routines and local goals seem to be on a different level than local choice. They have a lot to do with more global considerations about how everything fits together. Assembly lines have to be planned and laid out. Orchestral conductors need to make global choices about tempo and expressiveness. If these are not good then everyone can play their part perfectly but the overall product will be imperfect. Even good cooks, using good ingredients produce bad food if their recipe is wanting.

To study coordination at this level requires methods we have not discussed: modeling and simulation, scheduling theory, and others. Economic and computational models are of use in showing us the impact of different global coordinating mechanisms when we have actual systems to compare. If we do not have living versions of different systems of coordination, how can we predict the value of re-engineering a process? Only by modeling and simulating can we study the temporal effects of such things as changing the time and destination of resources, or the impact of changing the connectivity, reliability or speed of communication, or the pattern of messaging. Only through simulation can we begin to see how one participant's local activity in his own activity space can have side effects on neighboring or intersecting activity spaces and so produce a cascade of side effects.

For these reasons, modeling and simulation are necessary for exploring coordination and the robustness of a distributed system. But I'd like to close with two cautionary points concerning simulation and models. The first concerns the creativity which well motivated participants show in compensating for bad design or for torpedoing good design. The second concerns the importance of history.

In our Jolt's study baristas work in such tight quarters that their activity spaces physically overlap. This means that one person may put an item in the area another was intentionally keeping clear. With practice, baristas who share space get to know each others' work patterns and the two develop a new and better dynamic of coordination. This learning is idiosyncratic and hard to simulate.

Even more spectacular is the way people discover or learn to compensate for their own and their team's limitations. People with advanced Parkinson's disease, for example, lose the ability to control walking because of corruption in neural processing of proprioceptive input. Yet, if they walk on floors that have large stripes or checkerboard tiling they can compensate for the loss of proprioception by using the rhythmic input the visual system provides. How could such successes be predicted by models?

History is important because coordination in ongoing systems is almost always history dependent. To appreciate how hard adding history makes the problem of coordination, imagine that we set out to model and then simulate a distributed system in which agents rely on a clock to coordinate timing. Under reasonable assumptions we may be able to show that without a clock timing would be unacceptably bad. Great result! But where did our reasonable assumptions come from? Presumably from an idealization about the way the system in question operates right now. Yet if we have learned anything from looking at the complexity of systems, it is that evolution can find multiple paths to the same goal. For a large class of systems, including our target system, a differently designed system which relies on, say, conveyor belts moving at fixed speed covering a fixed distance can be temporally coordinated as well as any system with a clock. It depends on what needs to be where and when. This diversity of solutions highlights the need to stay close to the facts. We can never understand the elements driving the coordination of a natural distributed system if we suppose that the system, its setup, its timing, its rules and culture of operation, are devoid of history. Parts have been adapted at every level, and the form they are in now are a partial function of the form they were in before. If it were not so hard to know the aspects of a system that transfer well to the real world, business models would be more successful.

The upshot is that as designers we must always work from the present, mindful of the inertia of users. If we create a design that is too distant from current activities, however cognitively efficient we think it is, users will either not adopt it — so it is de facto ineffective — or users will co-opt it for their own purposes. The gulf between the theories we have and the designs we need remains wide.

6. Conclusion

In the search for a comprehensive view of how humans structure and interact with their environments, distributed cognition adds a meaningful constraint on the final story. What makes an environment a good place in which to work? Cognitive engineering would be considerably easier if we had a principled answer to this question. Designers could then consult the book of cognitive and interactive principles and, like civil engineers designing bridges, they could creatively apply principles to build environments that make life easier. At present, no such book exists. And the prospects of ever writing one depends on our understanding of a collection of issues broadly related to interaction and distributed cognition which we are just beginning to appreciate. In this paper, I have presented a few of these issues. Chiefly my focus has been on methods of measurement: economic, computational and conformity to principle. The complexity of real-world distributed cognition systems is so great and so dependent on details of history, personal expertise and physical layout that our models and metrics are bound to be inadequate. But we also are at a point in science where new mathematics, new modeling possibilities, and more extended videographic analyses will let us enter into the next level of measurement and observation. It is a good time to be a cognitive scientist.

Note

* It is a pleasure to thank Stevan Harnad and Sam Harvey.

References

Hutchins, E. 1995. *Cognition in the Wild*. Cambridge, MA: The MIT Press.

Kirsh, D. 1995. "The Intelligent Use of Space". *Artificial Intelligence* 73(1–2): 31–68.

Kirsh, D. 1996. "Adapting the environment instead of oneself". *Adaptive Behavior* 4(3–4): 415–452.

Kirsh, D. 1999. "Distributed cognition, coordination and environment design". In *Proceedings of the European conference on Cognitive Science*, 1–11.

Kirsh, D. 2001. "The context of work". *Human Computer Interaction* 16(2–4): 305–322.

Kirsh, D. 2004. "Metacognition, distributed cognition and visual design". In P. Gardinfors and P. Johansson (eds), *Cognition, Education and Communication Technology*. Hillsdale, NJ: Lawrence Erlbaum, 147–179.

Kirsh, D. and Maglio, P. 1994. "On distinguishing epistemic from pragmatic actions". *Cognitive Science* 18: 513–549.

McNeill D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: The University of Chicago Press.

Vicente, K. 2004. *The Human Factor: Revolutionizing the Way People Live with Technology*. London: Taylor and Francis.

Author's address

David Kirsh
Department of Cognitive Science
University of California San Diego
La Jolla, CA 92093-0515
U.S.A.

kirsh@ucsd.edu
http://adrenaline.ucsd.edu/kirsh

About the author

David Kirsh is Professor in the Cognitive Science Department at the University of California San Diego, where he has been since its inception in 1989. His D.Phil is from Oxford University in Philosophy and he held a post doc and later was research faculty at MIT in the Artificial Intelligence Lab. His interests are broad but recently concentrate on understanding principles for designing interactive environments.

Radical changes in cognitive process due to technology

A jaundiced view*

Arthur M. Glenberg
University of Wisconsin, Madison

A strong case can be made that the cognitive system is designed for guiding action, not, for example, symbol manipulation. I review empirical work demonstrating the link between action and cognition with special attention to the processes of language comprehension. Next, I sketch an embodied cognition framework for integrating work on language understanding with a more general approach to cognition and action. This general approach considers contributions to action of bodily states, emotions, social and cultural processes, and learning within a framework that generates a dynamic system. This framework is used to consider the notion of distributed cognition and the prospects that technology might induce substantial changes in cognition. My assessment is that such changes are unlikely.

Keywords: embodied cognition, language comprehension, distributed cognition, action

1. Grounding cognition in action

What are brains for? The biologist Rudolfo Llinas makes the answer very clear when he notes, "The nervous system is only necessary for multi-cellular creatures…that can orchestrate and express active movement" (Llinas 2001: 15). That is, brains are for the control of action. Llinas also describes a fascinating example of this principle: the sea squirt. The sea squirt begins life as a tadpole-like creature with a primitive spinal chord, nervous system, and sensory organs such as an eye spot. It swims until it finds a suitably stable place to which to attach. After attaching, it never again orchestrates or expresses active movement; and after attaching, it begins to digest its nervous system. No action, no brains.

Pragmatics & Cognition 14:2 (2006), 263–274.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

The idea that brains are for action can be used as an organizing principle to understand not just the sea squirt, but virtually all of human cognition from perception to action to language and into the classroom. It can also be used to provide a perspective on distributed cognition. My plan is to briefly review data showing the role of action in cognition (language comprehension, in particular), to sketch a theoretical framework based on this principle, and then to use data and theory to discuss the role of technological advances, such as the Internet, in changing human thinking.

1.1 Symbol grounding

One argument for action as the basis of cognition begins with the symbol grounding problem (Harnad 1990; Searle 1980). Some scientists have characterized human cognition as manipulation of abstract, amodal, and arbitrary symbols, much like words that are arbitrarily linked to their meaning or the patterns of bits in a computer that are arbitrarily representations of letters. But others have noted that human cognition also deals with meaningful content. Although both ideas seem right, they are in conflict: meaning cannot arise solely from the manipulation of arbitrary symbols. Harnad's symbol merry-go-round argument provides a good illustration of the problem. Suppose that you just arrived in a country and you don't know the local language. At your disposal, however, is a dictionary written in that language. You see what appears to be an informational sign, and you start to use your dictionary to try to understand the words on the sign. You look up the first word, but it is defined solely in terms of other words (abstract, amodal, arbitrary symbols) of the language you do not know. Undaunted, you look up the first word in the definition to find that it, too, is defined in terms of other unknown symbols. The point is the following: No matter how many words you look up, that is, no matter how many symbols you manipulate, you will never discover the meaning of any of the words, let alone the meaning of the sign. The abstract, amodal, and arbitrary symbols of language, mathematics, and computers must be grounded in something outside of the symbol system to generate meaning.

1.2 Grounding language in action

I (Glenberg 1997) and others have proposed that action is an appropriate medium for grounding. We know that brains evolved for action. Given that evolution tends to be conservative, we might suppose that other systems, for example, language, are built on a base of action (Gallese and Lakoff (2005) call this 'neural exploitation'). The Indexical Hypothesis (Glenberg and Robertson 1999, 2000; Kaschak and Glenberg 2000) explains how the abstract, amodal, and arbitrary symbols of language become embodied, action-based meaning. Consider a sentence such as, "Art stood on the chair to change the light bulb". According to the IH, understanding begins by indexing or mapping words and phrases (e.g., "the chair") to objects in the environment or to perceptual (that is, analogical) symbols (Barsalou 1999). Next, affordances (Gibson 1979) are derived from the objects. Affordances are possibilities for interaction that are jointly determined by characteristics of the objects and characteristics of bodies. Thus, chairs can afford sitting on for humans, but not for elephants. Kitchen chairs afford standing on, or hiding under for children, but not adults. The third step in the IH is to mesh, or combine, the affordances as directed by syntax. For example, the syntax of the example sentence directs us to combine the affordances so that Art is on the chair rather than vice versa.

1.3 Empirical support for grounding meaning in action

Evidence for the grounding of language in action can be found in a variety of domains. Consider neuroscience first. Hauk, Johnsrude and Pulvermuller (2004) used fMRI to record brain activity while people listened to verbs such as "lick", "pick", and "kick". When listening to "lick", there was activation in the area of motor cortex used for controlling the mouth; for "pick", activation appeared in the area of motor cortex controlling the hand; and for "kick", activation occurred in motor cortex controlling the leg. Buccino et al. (2005) used TMS to probe pre-motor cortex while listening to sentences describing hand (e.g., "He sewed") or leg activity. They also observed that sentence content differentially activated motor areas of the brain as well as more peripheral musculature.

Behavioral work has also produced evidence for grounding in action. For example, Glenberg and Kaschak (2002) asked participants to judge the sensibility of sentences such as "Open the drawer" and "Open the plate". Some sentences described action toward the reader, and other sentences described action away from the reader (e.g., "Close the drawer"). Orthogonally, in one condition the "Yes" response was made by literally moving the hand away from the body to a response key, and in another condition, the response was made by moving the hand toward the body to the response key. Participants viewed a sentence by pushing the middle button on a three-button panel held in the lap. The buttons were oriented so that pressing the farthest button required moving the hand away from the body, and pressing the nearest button required moving the hand toward the body. In one condition, the "Yes" response was made by pressing the farthest button, and in the other condition, the "Yes" was made by moving the hand toward the body to the nearest button. If understanding the sentence requires the derivation and mesh of affordances (i.e., a simulation of action), then having understood a sentence implying action in a particular direction (e.g., away from the body) should prime responding in the same direction. That is exactly what was found for imperative sentences, sentences describing the transfer of concrete objects ("Courtney handed you the notebook"), and sentences describing the transfer of information ("You told Liz the story"). De Vega (in press) has even reported evidence that highly abstract counterfactual language is grounded in action.[1]

Developmental psychology has also produced strong evidence for the importance of action in cognition throughout development. For example, Sommerville et al. (2005) gave 3-month-old infants precocious experience with object manipulation by having them wear mittens to which objects would stick. This experience changed the way the infants observed objected-directed actions of adults. With slightly older children, Campos et al. (2000) have documented the enormous changes in cognition and social interaction that are produced by the accomplishment of self-locomotion. Smith (2005) investigated effects of action experience on categorization and language learning with three-year olds. She had toddlers play with an asymmetrical object shaped roughly like a barbell with one end notably smaller than the other. The object was referred to as a "wug". In one condition, the children played with the wug by waving it in one hand while holding the less-bulbous side, thus emphasizing through action the asymmetry. In the other condition, the children used two hands to grasp the ends of the object symmetrically and rotate it. Children were then asked if members of a series of increasingly more symmetric objects were wugs and if members of a series of increasingly less symmetric objects were wugs. Children's acceptance of objects as members of the wug category depended on the type of actions they performed with the objects. Namely, children who used the wug in an asymmetric fashion were more likely to accept the new, more asymmetric objects as members of the wug category, and they were less likely to accept the more symmetric objects as members of the category. Finally, Glenberg et al. (2004) tested the hypothesis that some children are poor readers because they have never learned to index written words to the objects the words represent. For such children, reading is a boring exercise in word-calling rather than a meaning-producing activity. In their experiments, six- and seven-year old children read about activities on a farm and had in front of them a toy farm with a barn, tractor, etc. As children read each sentence, they manipulated the toys to correspond to the meaning of the sentence. This simple procedure led to memory and story comprehension close to two standard deviations greater than in the control condition in which children read and reread the story. The point of this work with infants, toddlers, and children is that action is central to cognition throughout cognitive development.

2. A general framework for cognition and action

The Indexical Hypothesis can be combined with Gallese and Lakoff's (2005) idea of neural exploitation and the organizing principle that brains are for action. Begin with the proposition that the basic function of the cognitive system is to select the next action, be it turning the head, taking a step, bestowing a kiss, reading a book, or investing a fortune in an internet start-up (see Figure 1). Action is selected by a process of constraint satisfaction. One source of constraints comes from the body (e.g., morphology and physiology) and the environment acting together to produce affordances. Other constraints come from learning and perceptual learning which control the derivation of affordances (Goldstone 2003).

The social situation is another potent source of constraints. A conspecific's actions both change the environment and change one's ability to act in that environment. Furthermore, primates have evolved mirror neuron systems that allow for the automatic understanding of another's actions (Rizzolatti and

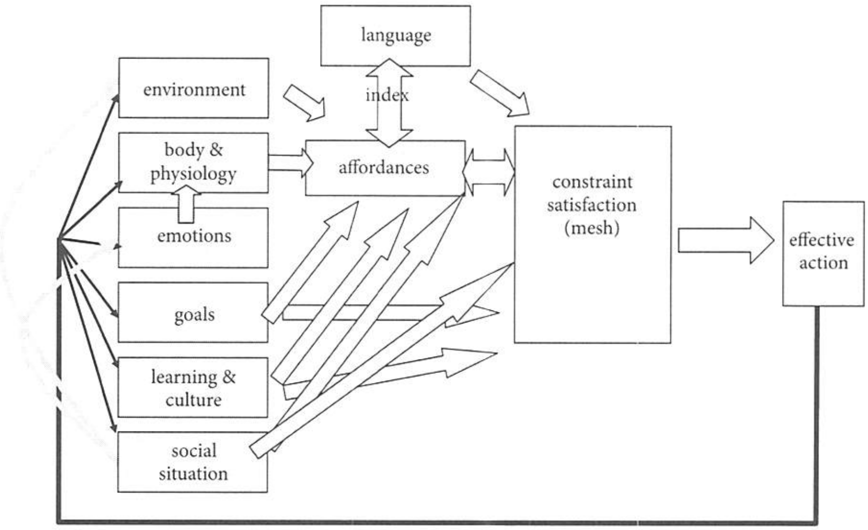


Figure 1. A functional model of bodily states, language, and action.

Craighero 2004; Fogassi et al. 2005). Mirror neurons in the pre-motor and parietal areas are activated both by taking action and by observing others taking the same action. This system allows us to recognize the intent of another's actions because it is the same intent that that we have upon taking that action. There is also an emotion mirror neuron system (Gallese et al. 2004). The operation of this system underlies our ability to empathize and understand the emotions of another, as well as to act accordingly.

As illustrated in Figure 1, once an action is generated, that action will itself change the environment, the social situation, and engender learning. Thus, taking one action changes the constraints for selecting the next action and in the process creates a dynamic system (e.g., van Orden et al. 2003).

In accord with the principle of neural exploitation (Gallese and Lakoff 2005), language takes advantage of the system originally designed by evolution to produce action. Thus, the affordances that are used in language comprehension are the same affordances that are used in constraining actions. Similarly, the process of mesh guided by syntax is the same process of constraint satisfaction otherwise guided by goals, social constraints, and so on. On this account, language is not a separate module. Instead, language comprehension utilizes the same neural machinery used in guiding all actions, and language comprehension is affected by the same considerations of body, emotions, social situations, and so forth.

3. Action framework and distributed cognition

Hollan et al. (2000) offer three characteristics of distributed cognition. First, cognitive processes may be distributed across the members of a social group. I understand this to mean that different members of a group will provide different types of expertise to joint activity, and hence the final product is likely to be quite different from what would be produced in isolation. Second, cognitive processes may involve coordination between internal and external structures. Third, cognitive processes may be distributed through time so that the final product reflects earlier products.

Given these features, the framework illustrated in Figure 1 can be characterized as an example of distributed cognition. First, members of the social group affect the constraints that go into selecting an action through (a) their direct effect on the environment, (b) our understanding of their action using mirror neurons, (c) our response to their emotional states (again using mirror neurons), and (d) their ability to drive our constraint satisfaction system through language. Then, given how our own responses will effect the environment as well as affecting others in the environment, the result is a complex, dynamic system of interlocking actions that will be driven in ways that neither partner can confidently predict. Second, the system illustrated in Figure 1 uses affordances as a primary source of constraints on thinking. Affordances arise from the physical situation (or memory based on previous interactions with the physical system) and the body. Thus, like distributed cognitive systems, the framework in Figure 1 involves the coordination of internal and external structures. Finally, given the feedback system illustrated by the arrows in the figure, as well as the effects of action on learning, the framework is consistent with the claim that cognition is distributed in time.

According to Harnad (2005), the system described in Figure 1 and above is more properly characterized as generating "collaborative cognition" rather than "distributed cognition". Harnad wishes to reserve the latter term for a system that not only thinks but also can feel itself thinking. I do not wish to enter that debate. For me, the important point is to recognize that thinking is tremendously affected by the physical and social situations in which it is embedded so that a proper account of thinking requires a theoretical account of those situations and their impact on the cognitive system.

4. Implications for technology

Given the collaborative/distributed system in Figure 1, what are the implications for technology? Will the internet "induce substantive changes in any aspect of cognition itself" (Harnad and Dror, *Pragmatics & Cognition*, Call for Papers for the Distributed Cognition Special Issue)? Will the internet and other technologies "shift from being mere tools that aid cognition to having constitutive roles in shaping cognitive processes themselves" (Dascal and Dror 2005: 452)? Will investigations of e-learning and e-training "prove to be groundbreaking for the understanding of the crucial cognitive process of acquiring and using knowledge" (Dascal and Dror 2005: 454)? In brief, I am skeptical.

One reason for skepticism is that even evolutionary changes in cognition appear to be incremental, not radical. For example, although cognitive systems have certainly advanced enormously from the sea squirt, it is still a good bet that the human cognitive system serves primarily to control action. Furthermore, given that action must respect constraints of the body, whether the likes of a human body or a sea squirt body, basic principles of cognition likely apply to all animals (see, for example, von Uexküll's (1957) discussion of the cognitive world of the tick). Perhaps technology can forge changes in cognition that hundreds of millions of years of evolution have not, but that is not a good bet.

A second reason for skepticism arises from the framework in Figure 1. To the extent that the framework is correct, any benefits due to technology need to shape themselves to a human cognitive system based on action, rather than trying to re-program the cognitive system as if it were a general-purpose computer. Hollan et al. (2000) call for just these sorts of efforts, and there have been a few notable efforts to shape technology in this way. For example, Gee (2003) has put educational material into a gaming environment to take advantage of the motivating power of action games. Another example is *3D Reader* (Johnson-Glenberg, in press), which uses narrative texts to illustrate scientific principles and requires readers to actively create visual models of the principles. But it is an open question whether these attempts can be successfully generalized.

A third reason for skepticism is that most e-learning and e-training remove from the situation just those aspects that the human cognitive system has evolved to take advantage of. For example, e-learning systems use few real manipulables. By real manipulables, I mean objects that have affordances relevant to the situation at hand. For example, in the experiments reported in Glenberg et al. (2004) children manipulated toy animals, tractors, and so forth, to correspond to the actions in the stories they were reading about a farm. Clearly, a toy tractor does not have the affordances of a real tractor. But the toys do have affordances that mesh with other affordances in the toy farm context. For example, the toy tractor is the right size to pull the toy cart, and the toy cart can carry the toy cow. Because the toys have the right affordances, they provide the right constraints for understanding the texts. Glenberg et al. (in press) demonstrated that the right affordances can also be important for understanding in mathematics. In those experiments, children in the third grade read and solved mathematical story problems based on a toy zoo. In one condition, children manipulated and counted story-relevant objects (e.g., toy fish to use as food) as they read. In another condition, children were given "abstract" manipulables, that is, Legos that they could manipulate and count out but had few of the affordances of the objects they were reading about. Children who manipulated story-relevant objects increased their problem solving by 1.5 standard deviations over the children using the abstract manipulables.

Another component removed from most e-learning situations is real human interaction. Human communication often involves more than words. For example, in face-to-face learning environments, teachers' hand and arm gestures convey a tremendous amount of information that students use (Valenzeno et al. 2003). Another source of information in face-to-face communication consists of emotions. As noted before, people have evolved mechanisms for picking up and responding to emotions conveyed by others (Gallese et al. 2004). Furthermore, experiencing an emotion can play an important role in language comprehension (Glenberg et al. 2005). Finally, real interaction allows for the establishment of common ground (Clark 1996; Pickering and Garrod 2004), that is, a set of objects, actions, and experiences referred to in the conversation, established ways of referring to those entities, and even jointly-determined syntactic structures that ease communication and facilitate understanding. Consequently, even if the e-learning involves multi-media presentations, without a teacher who can jointly ground the language with the student, who can gesture appropriately to convey information difficult to convey in words, who can portray excitement, happiness, and fear, it is hard to imagine that e-learning will be anything more than a pale substitute for real teachers and real human interaction (but see Graesser and Jackson, in press, for a discussion of incorporating emotional processes into an automated tutor).

A final reason for skepticism concerns temporal limits on human learning. To my knowledge, there is little work on the question of how much can be learned per unit of time, but my guess is that the answer will be some variation of "not much". We can only read so fast, listen so fast, and see so fast; if we try to go faster, the signal becomes an unintelligible blur. Furthermore, even when we succeed in learning material in one context, often there is little transfer to related contexts (Detterman 1993; Gick and Holyoak 1980, 1983; but see Goldstone et al., in press). Consequently, we must repeatedly practice the application of the same principles across minor variations in contexts, further slowing the rate at which information can become useful. Thus, although technology may offer a limitless smorgasbord of learning opportunities, our cognitive system places limits on how much can be usefully digested.

My view of technology and cognition is not completely jaundiced. I am as addicted to email as any other academic, and Google Scholar has reduced my trips to the university's library to almost none. A positive consequence is that I have more worldwide contacts than I would have imagined possible a decade ago. Also, I read more widely, although probably not more because of limits on time. But, are my thinking processes qualitatively different from a decade ago? I don't detect a difference. Nonetheless, to end on a positive note, in the spirit of Dascal and Dror (2005) and Hollan et al. (2000), one can view most of my reasons for skepticism as challenges for the future development of technology that is sensitive to the principles of biological cognitive systems.

Notes

* This work was supported by NSF grant BCS-0315434 and Institute of Education Sciences grant R305H030266 to Arthur Glenberg. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the funding agencies.

1. The emphasis on action does not imply that other bodily systems play no role in language. For example, there is also strong evidence that language calls on perceptual systems (see Kaschak et al. 2005). According to the IH, these perceptual systems are tapped whenever one has to derive affordances and conduct simulations. For example, in reading a sentence describing visual motion, one uses the visual system, area V5/MT in particular (Tetamanti et al. 2005), to simulate visual motion.

References

Barsalou, L.W. 1999. "Perceptual symbol systems". *Behavioral and Brain Sciences* 22: 577–660.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G.. 2005. "Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study". *Cognitive Brain Research* 24: 355–363.

Campos, J.J., Anderson, D.I., Barbu-Roth, M.A., Hubbard, E.M., Hertenstein, M.J., and Witherington, D. 2000. "Travel broadens the mind". *Infancy* 1: 149–219.

Clark, H.H. 1996. *Using language*. Cambridge: Cambridge University Press.

Dascal, M. and Dror, I.E. 2005. "The impact of cognitive technologies: Towards a pragmatic approach". *Pragmatics & Cognition* 13: 451–457.

Detterman, D.R. 1993. "The case for prosecution: Transfer as an epiphenomenon". In D.K Detterman and R.J Sternberg (eds), *Transfer on Trial: Intelligence, Cognition, and Instruction*. Norwood, NJ: Ablex Publishing Corporation, 1–24.

De Vega, M. In press. "Levels of embodied meaning: From pointing to counterfactuals". In Glenberg et al. (eds).

Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. 2005. "Parietal lobe: From action organization to intention understanding". *Science* 308: 662–667.

Gallese, V., Keysers, C, and Rizzolatti, G. 2004. "A unifying view of the basis of social cognition". *Trends in Cognitive Science* 8: 396–403.

Gallese, V. and Lakoff, G. 2005. "The brain's concepts: The role of the sensory-motor system in conceptual knowledge". *Cognitive Neuropsychology* 22: 455–479.

Gee, J. 2003. *What Video Games Have to Teach us about Learning and Literacy*. New York: Palgrave-MacMillan.

Gibson, J.J. 1979. *The Ecological Approach to Visual Perception*. New York: Houghton Mifflin.

Gick, M.L. and Holyoak, K.J. 1980. "Analogical problem solving". *Cognitive Psychology* 12: 306–355.

Gick, M.L. and Holyoak, K. J. 1983. "Schema induction and analogical transfer". *Cognitive Psychology* 15: 1–39.

Glenberg, A.M. 1997. "What memory is for". *Behavioral and Brain Sciences* 20: 1–55.

Glenberg, A.M., Gutierrez, T., Levin, J.R., Japuntich, S., and Kaschak, M.P. 2004. "Activity and imagined activity can enhance young children's reading comprehension". *Journal of Educational Psychology* 96: 424–436.

Glenberg, A.M., Havas, D.A., Becker, R., and Rinck, M. 2005. "Grounding language in bodily states: The case for emotion". In R. Zwaan and D. Pecher (eds). *The Grounding of Cognition: The Role of Perception and Action in Memory, Language, and Thinking.* Cambridge: Cambridge University Press, 115–128.

Glenberg, A.M., Jaworski, B., Rischal, M, and Levin, J.R. In press. "What brains are for: Action, meaning, and reading comprehension". In D. McNamara (ed), *Reading Comprehension Strategies: Theories, Interventions, and Technologies.* Mahwah, NJ: Lawrence Erlbaum.

Glenberg, A.M. and Kaschak, M.P. 2002. "Grounding language in action". *Psychonomic Bulletin & Review* 9: 558–565.

Glenberg, A.M., and Robertson, D.A. 1999. "Indexical understanding of instructions". *Discourse Processes* 28: 1–26.

Glenberg, A.M. and Robertson, D.A. 2000. "Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning". *Journal of Memory and Language* 43: 379–401.

Glenberg, A.M., de Vega, M., and Graesser, A.C. (eds). In press. *The Garachico Workshop on Symbols and Embodiment.* Oxford: Oxford University Press.

Goldstone, R.L. 2003. "Learning to perceive while perceiving to learn". In R. Kimchi, M. Behrmann, and C. Olson (eds). *Perceptual Organization in Vision: Behavioral and Neural Perspectives.* Hillsdale, NJ: Lawrence Erlbaum, 233–278.

Goldstone, R.L., Landy, D.C., and Son, J.Y. In press. "A well grounded education: The role of perception in science and mathematics". In Glenberg et al. (eds).

Graesser, A.C. and Jackson, G.T. In press. "Body and symbol in autotutor: Conversations that are responsive to the learners' cognitive and emotional states". In Glenberg et al. (eds).

Hauk, O., Johnsrude, I., Pulvermüller, F. 2004. "Somatotopic representation of action words in human motor and pre-motor cortex". *Neuron* 41: 301–307.

Harnad, S. 1990. "The symbol grounding problem". *Physica D* 42: 335–346.

Harnad, S. 2005. "Distributed processes, distributed cognizers and collaborative cognition". *Pragmatics & Cognition* 13(3): 501–514.

Hollan, J., Hutchins, E., and Kirsh, D. 2000. "Distributed cognition: Toward a new foundation for human–computer interaction research". *ACM Transactions on Computer–Human Interaction* 7: 174–196.

Johnson-Glenberg, M.C. In press. "Web-based reading comprehension instruction: Three studies of 3D-Readers". In D. McNamara (ed), *Reading Comprehension Strategies: Theories, Interventions, and Technologies.* Mahwah, NJ: Lawrence Erlbaum.

Kaschak, M.P. and Glenberg, A.M. 2000. "Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension". *Journal of Memory and Language* 43: 508–529.

Kaschak, M.P., Madden, C.J., Therriault, D.J., Yaxley, R.H., Aveyard, M., Blanchard, A., and Zwaan, R.A. 2005. "Perception of motion affects language processing". *Cognition* 94: B79-B89.

van Orden, G.C., Holden, J.G., and Turvey, M.T. 2003. "Self-organization of cognitive performance". *Journal of Experimental Psychology: General* 132: 331–350.

Llinas, R. 2001. *The i of the Vortex*. Cambridge, MA: The MIT Press

Pickering, M.J. and Garrod, S. 2004. "Toward a mechanistic psychology of dialogue". *Behavioral and Brain Sciences* 27: 169–225.

Rizzolatti, G. and Craighero, L. 2004. "The mirror-neuron system". *Annual Review of Neuroscience* 27: 169–192.

Searle, J.R. 1980. "Minds, brains, and programs". *Behavioral and Brain Sciences* 3: 417–424.

Smith, L. B. 2005. "Action alters shape categories". *Cognitive Science* 29: 665–679.

Sommerville, J.A., Woodward, A.L., and Needham, A. 2005. "Action experience alters 3-month-old infants' perception of others' action". *Cognition* 96: B1-B11.

Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S.F., and Perani, D. 2005. "Listening to action-related sentences activates fronto-parietal motor circuits". *Journal of Cognitive Neuroscience* 17: 273–281.

Valenzeno, L., Alibali, M.W, Klatzky R. 2003. "Teachers' gestures facilitate students' learning: A lesson in symmetry". *Contemporary Educational Pscyhology* 28: 187–204.

von Uexküll, J. 1957. "A strol through the worlds of animals and men — a picture book of invisible worlds". In C.H. Schiller (ed), *Instinctive Behavior — The Development of a Modern Concept*. New York: International Universities Press, 5–80

Author's address

Arthur Glenberg
Department of Psychology
1202 West Johnson Street
Madison, WI 53706
U.S.A.

Glenberg@wisc.edu

About the author

Arthur Glenberg is professor of Psychology and Educational Psychology at the University of Wisconsin. He has held visiting scientist positions at the University of Colorado, the Institute for Interdisciplinary Studies (ZiF), the Max Planck Institute for Cognitive Science and Neuroscience, and the Department of Neuroscience in Parma, Italy. His research is developing a theory of language comprehension based on the embodied cognition framework.

The grounding and sharing of symbols

Angelo Cangelosi
University of Plymouth

The double function of language, as a social/communicative means, and as an individual/cognitive capability, derives from its fundamental property that allows us to internally re-represent the world we live in. This is possible through the mechanism of symbol grounding, i.e., the ability to associate entities and states in the external and internal world with internal categorical representations. The symbol grounding mechanism, as language, has both an individual and a social component. The individual component, called the "Physical Symbol Grounding", refers to the ability of each individual to create an intrinsic link between world entities and internal categorical representations. The social component, called "Social Symbol Grounding", refers to the collective negotiation for the selection of shared symbols (words) and their grounded meanings. The paper discusses these two aspects of symbol grounding in relation to distributed cognition, using examples from cognitive modeling research on grounded agents and robots.

Keywords: symbol grounding, cognitive modeling, neural networks, artificial life

1. Introduction

The main and most obvious function of language is its pragmatic role, i.e., that of facilitating interaction and communication between cognitive agents. Individuals use language to provide information (representative speech acts), ask for something (directive acts) and express internal states and emotions (expressive acts). A group of individuals engaged in dialog can be considered as a "dynamical" distributed cognitive system, where the different speakers (and listeners) collaborative perform a distributed social and cognitive task. Communication can be achieved not only through speaking, but also through other more "stable" forms of communication, such as written texts, audio recordings or other social artifacts. The fact that language is the means to represent and

Pragmatics & Cognition 14:2 (2006), 275–285.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

communicate about the world has one important consequence. It can also be considered as an autonomous distributed cognitive system, partially independent from the individual cognitive agents that have initially produced it. The system is distributed not only in the mind of language speaking individuals, but also in all other expressions of language such as books and the world wide web. It has also been suggested that this system can follow its own evolutionary dynamics, through the mechanisms of cultural transmission and meme evolution (Blackmore 1999; Croft 2000).

Language can also play a purely social and emotional role, when its communicative function is minimal or can sometime become totally irrelevant. For example, language and communication can be used as a means to increase bonding between the members of a family or social group. It has been suggested that the evolutionary origins of language are linked to that of grooming behavior in animal groups (Dunbar 1996).

In addition to being a social and pragmatic phenomenon, language has important "solipsistic" and cognitive functions. Vygotsky (1962, 1978) highlighted the importance of "speaking to oneself" as an essential cognitive mechanism in the development of the mind, although in an important social context. For example, recent computational modeling studies have demonstrated the role of speaking to oneself as an aid to cognition and categorization (Mirolli and Parisi 2005a; 2005b). The debate on linguistic relativism (Gumperz and Levinson 1996; Gentner and Goldin-Meadow 2003; also known as Whorf-Sapir Hypothesis) has produced much evidence that the structure and lexicon of language influence how we perceive and conceptualize the world. This has also cultural implications, when we consider the effects of different languages in the way peoples represent and describe their world. For example, in the case of spatial language, different languages describe and carve the space in quite different ways (Coventry and Garrod 2004; Bowerman et al. 2001).

The double function of language, as a social/communicative means, and as an individual/cognitive capability, derives from its fundamental property that allows us to internally re-represent the world we live in, to process such representations to generate new concepts and explanations, and to communicate with others about our knowledge of the world. This is possible through the mechanism of symbol grounding, i.e., the ability to associate entities and states in the external (and internal) world with internal categorical representations. These representations become symbolic when we are able to combine them to generate sentences describing new concepts and meanings. The ability to create categories, such as the categorical perception mechanisms, and to combine their symbolic representations to generate new meanings, have been proposed by Harnad (1987) as the groundwork of cognition. This has also been invoked in explaining the adaptive value of language and its evolutionary origins (Cangelosi and Harnad 2000).

The symbol grounding mechanism, as language itself, has both an individual and a social component. The individual component, called the "Physical Symbol Grounding", refers to the ability of each individual to create an intrinsic link between world entities and internal categorical representations. The social component, called "Social Symbol Grounding", refers to the collective negotiation for the selection of shared symbols (words) and their grounded meanings. In the next two sections we will analyze the two aspects of symbol grounding and their contribution to distributed cognition. We will use examples from computational cognitive modeling research to demonstrate how to design cognitive systems able to develop linguistic capabilities of physical and social symbol grounding. Our understanding of the functioning of artificial cognitive systems will, in turn, improve our knowledge of the organization of natural cognitive systems.

2. Physical symbol grounding

The symbol grounding problem (Harnad 1990) refers to the need, for natural and artificial cognitive agents, to acquire an intrinsic link between symbolic representations and referents in the external word. This is possible through the process of categorization that allows an individual to create an internal categorial representation of the external referent, and use it as the grounding (meaning) for symbols. These internal representations are acquired during interaction with entities in the external world. During category learning, we can "sort out" our environment by forming discrete, useful categories. The cognitive mechanism at the basis of the physical symbol grounding is based on the following chain of entities and representations:

external entities ⇔ internal representations ⇔ symbols

These are all bi-directional links that permit the external entities to influence internal representation and symbols, and at the same time the symbols to affect the way we represent our external world (Roy, in press).

The physical symbol grounding problem consists in the individual's cognitive ability to create such links. The social symbol grounding problem, as discussed in the next section, consists in the need to acquire a set of symbols. These are used by a group of agents to refer consistently to the same external referents, or to some of their properties. Although Harnad (1990) does not explicitly restrict the use of the term symbol grounding to an individual phenomenon, his discussion mostly focuses on the individual's cognitive capability for grounding symbols. Harnad's proposal that the internal representations that mediate grounding are essentially categorical has some important cognitive consequences. One of the fundamental characteristics of these representations is that of categorial perception (Harnad 1987). This is the process of warping the similarity space of internal categorial representations during learning. The perceptual, iconic representations of the members of a category are transformed so that the internal within-category differences are reduced (within-category compression). The differences between the iconic representations of the members of different categories are enhanced to make them appear more dissimilar (between-category expansion). These warped representations of the word facilitate the production of discrete representations that, as symbols, can be easily combined into propositions using logical and syntactic constructs.

Physical symbol grounding is a fundamental capability of natural (human) cognitive systems, since we are evolutionarily endowed with the competence to build categorical representations of the world. Categorical perception is widespread in natural cognitive systems, and has been observed in animals (e.g., Zentall et al. 1986) and humans (e.g., Goldstone 1994). In artificial cognitive systems research, however, physical symbol grounding can be a real "problem". In computational cognitive models purely based on symbolic representations and rule-based systems, agents use ungrounded symbols that require the interpretation of an external user (the researcher) to "understand" the meaning associated to symbols. Even if these models may include an additional set of symbols for the "meanings", such as a dictionary, these would still be ungrounded and self-referential symbols. Such situation is similar to the Chinese room scenario (Searle 1980) where the symbolic task of responding to questions in Chinese, without knowing the language, can apparently be well performed without understanding the meaning of the questions and answers.

Various approaches have been proposed to build grounded cognitive models based on physical symbol grounding. Hybrid symbolic-connectionist models have been proposed as ideal candidates for solving the symbol grounding problem (Harnad 1993; Sun 2002). The connectionist, neural network component is used to perform the task of grounding the symbols into perceptual and categorial representations. Input units receive iconic representation of the objects. These are transformed, in the hidden units, into warped, categorial perception representations. A symbolic module, such as a rule-based system, would use these categorial representations to perform symbol manipulation operations and high-level cognitive tasks (Miikkulainen 1994). An alternative approach to hybrid systems is that solely based on connectionist architectures (Cangelosi 2005; Cangelosi, Greco, and Harnad 2000; Harnad, Hanson, and Lubin 1994). These typically consist of neural networks that are able to acquire categorial perception representations during category learning tasks, to use these representations to ground the meaning of discrete symbols, and to perform simple symbol manipulation tasks.

A more recent approach to the symbol grounding problem in artificial cognitive systems relies on the use of embodied agent models and cognitive robotics (Vogt 2002; Cangelosi and Riga, in press; Cangelosi, in press). These methodologies permit the grounding of symbols directly in sensorimotor representations, such as action categories. Some of these robotic approaches also include the use of connectionist networks, while others use different control architectures. This robotic approach, when embedded in a population-based setting, also permits the simulation of the social symbol grounding process, as discussed below.

3. Social symbol grounding

Social symbol grounding refers to the process of developing a shared lexicon of perceptually-grounded symbols in a population of cognitive agents. In natural cognitive systems, the social grounding of symbols can be observed from different temporal perspectives. In slow, evolutionary terms, it refers to the gradual emergence of language. Our ancestors started from a pre-linguistic, animal-like society with no explicit symbolic and communicative means. During evolution, this led to the collective development of shared languages used to talk about entities in the physical, internal and social world. In ontogenetic terms, social symbol grounding refers to the process of language acquisition and cultural transmission. In early age, children acquire the language of the groups they belong to via imitation of their parents and peers. This leads to the gradual discovery and construction of linguistic knowledge (Tomasello 2003). During adulthood, this process continues through the general mechanisms of cultural transmission. Phenomena of cultural transmission such as variation, competition and preferential selection of words can result in slow historical changes in the language being transmitted (Croft 2000; Labov 2000). Finally, social symbol grounding can also happen during temporary interactional games between individuals as in interactive alignment in dialogue. During dialogue, interlocutors fixate local interpretations for expressions (routinization).

This helps them to converge on particular ways of using the language in specific contexts (Garrod and Anderson 1987; Pickering and Garrod 2004).

Social symbol grounding involves not only speech, but also a more distributed approach to communication and symbol use. This entails a wider interpretation of communication symbols that includes artifacts and cultural symbols. For example, in evolutionary and ontogenetic terms, the invention of writing has had an important role on language evolution and transmission. In dialogue contexts, it has been demonstrated that individuals are capable of negotiating and developing shared on-line sets of graphical conventions (Fay 2004). Such a distributed view of language leads to a fully distributed cognition system, with symbols existing in the mind of the language-speaking agents, in other cultural artifacts and symbolic entities, and in the general distributed cognition capabilities of the individuals. The act of communication in a group of individuals constitutes a collaborative cognitive task based on negotiation and dialogue-based interactions. Following this distributed view, language development can be seen as under dual control by adult and child (Cowley, in press). Both parties gear to each other's biomechanics and norm-based behavior prompting affective processes that drive prepared learning.

The issue of social symbol grounding has been directly addressed in various cognitive agent and robotics models of the emergence of language (e.g., Steels 1999; Cangelosi 2001; Cangelosi and Parisi 2002). These simulate the emergence of language in a group of cognitive agents that are able to ground and negotiate a shared set of communication symbols through direct negotiation. Steels and collaborators (1999) have developed a model of the cultural emergence of language in a population of robotic agents called "Talking Heads". These robots perform a collaborative task, based on language games, to name objects (colored shapes) randomly placed on a whiteboard. The speaker selects the topic of communication through shared attention, and produces an utterance to communicate its internal representation of the topic. In turn, the hearer guesses the meaning (topic) of the utterance and updates its linguistic knowledge. The success/failure of the language game results in the update of the agents' own lexicons, including the invention of new words when necessary. This model has been extended to study a web-based distributed cognitive system consisting of simulated agents and humans. Human participants can, through a web-based interface, be "embodied" into Talking Head agents. This allows them to play language games with other robots or humans, to learn the existing artificial lexicons, and also to influence them by inventing new words.

Other grounded agent models have simulated the emergence of language through the simulation of both evolutionary and ontogenetic learning processes. Cangelosi and collaborators (Cangelosi 2001; Cangelosi and Harnad 2000; Munroe and Cangelosi 2003) have simulated the emergence of compositional lexicons in a population of foraging agents in a mushroom world metaphor (Harnad 1987). Simulated agents evolve, through a genetic algorithm, according to their ability to recognize and name mushrooms. They must eat edible mushrooms and discard toadstools. During their lifetime, parent agents communicate with their children by producing two-word utterances describing actions on mushrooms. Child agents also learn their parents' lexicons through a process of imitation and cultural transmission. Although at the beginning of evolution the lexicon is totally random and meaningless, towards the end of the evolution the agents are able to evolve shared compositional languages, i.e., lexicons consisting of the name of an action ("approach" or "avoid") and the name of a mushroom (e.g., "champignon" or "porcino"). This model has been used to study the emergence of compositional languages based on two-word combinations. Simulations demonstrate that the structure of the lexicon directly reflects the sensorimotor structure of the stimuli (Cangelosi 2001; Cangelosi and Parisi 2004). The manipulation of cultural transmission mechanisms also affects the evolution of language through the Baldwin effect (Munroe and Cangelosi 2003).

4. Conclusion

The models described above demonstrate that the combination of social and physical grounding of symbols leads to the design of psychologically-plausible cognitive systems. Cognitive agents and robots are able to acquire a language based on a shared set of grounded symbols. These agents are also able to autonomously transfer the grounding of symbols from directly grounded words to higher-order symbols. At the beginning, they learn new symbols through trial-and-error and feedback-supervised experience with their perceptual referents. Once a set of grounded basic symbols is acquired, the agents can use these symbols to communicate with each other and create new meanings and new definitions of the world. For example, the combination of the basic symbols "horse" and "stripe" can be used to describe a new concept, that of "zebra". They can even create concepts of entities that might not necessarily exist in the world, such as the definition of "unicorn" through the combination of the basic and directly grounded concepts of "horse" and "horn". The combination of previously-grounded symbols such as "horse", "stripe" and "horn" permits the transfer of grounding from the basic-order symbols to new higher-order symbols. This has been demonstrated to occur in connectionist models of symbol grounding (Cangelosi et al. 2000) and in robotic models of the grounding and combination of action words (Cangelosi and Riga, in press).

The procedure used for the autonomous production of high-order symbols, as in Cangelosi and Riga's (in press) model of action name grounding, can be considered an implementation of Barsalou's (1999) perceptual symbol system hypothesis. In particular, when the agents, in a distributed and collaborative way, provide definitions of new concepts, they apply a symbol productivity mechanism. During word combination and the transfer of grounding, the agent plays some kind of internal re-enactment (simulation) of the meaning of the basic symbols. These re-activated internal representations are then used by the agent to merge the results of the two basic "mental" simulations and create a new combinatorial, grounded symbol and their corresponding combinatorial meaning (Wisniewski 1997).

The combination of the individual's physical grounding process and the group's social grounding of symbols permit the design of a psychologically-plausible distributed cognitive system. Knowledge is distributed in the individual agent's linguistic representation of the world, as well as in the social groups sharing the same symbols. Future research should consider extending this distributed approach to study additional phenomena in distributed cognition. For example, future grounded cognitive agent models might serve to investigate the emergence of physical and cognitive artifacts that agents use in collaborative and distributed tasks. Although some past agent-based models have focused on the emergence of artifacts (Ugolini and Parisi 1999), they have not addressed their explicit role in a distributed view of language (Cangelosi, in press; Coley, in press).

References

Barsalou, L. 1999. "Perceptual symbol systems". *Behavioral and Brain Sciences* 22: 577–609.

Blackmore, S. 1999. *The Meme Machine*. Oxford: Oxford University Press.

Bowerman, M., Levinson, S., and Levinson, S.C. 2001. *Language Acquisition and Conceptual Development*. Cambridge: Cambridge University press.

Cangelosi, A. 2001. "Evolution of communication and language using signals, symbols and words". *IEEE Transactions on Evolutionary Computation* 5(2): 93–101.

Cangelosi, A. 2005. "Approaches to grounding symbols in perceptual and sensorimotor categories". In H. Cohen and C. Lefebvre (eds), *Handbook of Categorization in Cognitive Science*. London: Elsevier, 719–737.

Cangelosi, A. In press. "Adaptive agent modeling of distributed language: Investigations on the effects of cultural variation and internal action representations". *Language Sciences*. Special issue on cognitive dymanics and distributed view of language.

Cangelosi, A., Greco, A., and Harnad, S. 2000. "From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories". *Connection Science* 12(2): 143–162.

Cangelosi, A. and Harnad, S. 2000. "The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories". *Evolution of Communication* 4(1): 117–142.

Cangelosi, A. and Parisi D. (eds) 2002. *Simulating the Evolution of Language*. London: Springer.

Cangelosi, A. and Parisi, D. 2004. "The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging". *Brain and Language* 89(2): 401–408.

Cangelosi, A. and Riga, T. In press. "An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots". *Cognitive Science*.

Cowley, S. In press. "Distributed language: Biomechanics, functions and the origins of talk". *Language Sciences*. Special issue on cognitive dymanics and distributed view of language.

Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow, Essex: Longman.

Dunbar, R. 1996. *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press

Fay, N. 2004. "The emergence and evolution of graphical productions". In *Proceedings of EELC1: First International Workshop on the Emergence and Evolution of Linguistic Communication*, Kanazawa, 53–58.

Garrod, S. and Anderson, A. 1987. "Saying what you mean in dialog: A study in conceptual and semantic co-ordination". *Cognition* 27: 181–218.

Gentner, D. and Goldin-Meadow, S. (eds). 2003. *Language in Mind*. Cambridge, MA: The MIT Press.

Gumperz, J.J., and Levinson, S.C. (eds). 1996. *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press.

Goldstone, R. 1994. "Influences of categorization of perceptual discrimination". *Journal of Experimental Psychology: General* 123: 178–200.

Harnad, S. 1990. "The symbol grounding problem". *Physica D* 42: 335–346.

Harnad, S. 1993. "Grounding symbols in the analog world with neural nets". *Think* 2(1): 12–78.

Harnad, S. (ed). 1987. *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press

Harnad, S., Hanson, S.J., and Lubin, J. 1994. "Learned categorical perception in neural nets: Implications for symbol grounding". In Honavar and Uhr (eds), 191–206.

Honavar, V. and Uhr, L. (eds). 1994. *Symbol Processors and Connectionist Network Models in Artificial Intelligence and Cognitive Modelling: Steps Toward Principled Integration*. New York: Academic Press.

Labov, W. 2000. *Principles of Linguistic Change*. Oxford: Blackwell.

Miikkulainen, R. 1994. "Integrated connectionist models: Building AI systems on subsymbolic foundations". In Honavar and Uhr (eds), 483–508.

Mirolli, M. and Parisi, D. 2005a. "How can we explain the emergence of a language which benefits the hearer but not the speaker?". *Connection Science* 17(3–4): 307–324.

Mirolli, M., and Parisi, D. 2005b. "Language as an aid to categorization: A neural network model of early language acquisition". In A. Cangelosi et al. (eds), *Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*. Singapore: World Scientific.

Munroe, S. and Cangelosi, A. 2002. "Learning and the evolution of language: The role of cultural variation and learning cost in the Baldwin Effect". *Artificial Life* 8: 311–339.

Pickering, M. and Garrod, S. 2004. "Toward a mechanistic psychology of dialogue". *Behavioral and Brain Sciences* 27: 169–226.

Roy, D. In press. "Semiotic Schemas: A framework for grounding language in action and perception". *Artificial Intelligence*.

Searle, J.R. 1980. "Minds, brains, and programs". *Behavioral and Brain Sciences* 3(3): 417–457.

Steels, L. 1999. *The Talking Heads Experiment. Volume 1. Words and Meanings*. Antwerpen: Laboratorium.

Sun, R. 2002. "Connectionist implementationalism and hybrid systems". In *Encyclopedia of Cognitive Science*. London: Nature Publishing Group — MacMillan, 697–703.

Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Ugolini, M. and Parisi, D. 1999. "Simulating the evolution of artifacts". In D. Floreano, J.-D. Nicoud, and F. Mondada (eds), *Advances in Artificial Life. ECAL99*. New York: Springer, 489–498.

Vogt, P. 2002. "The physical symbol grounding problem". *Cognitive Systems Research* 3(3): 429–457.

Wisniewski, E.J. 1997. "When concepts combine". *Psychonomic Bulletin and Review* 4(2): 167–183.

Zentall, T.R., Jackson-Smith, P., Jagielo, J.A., and Nallan, G.B. 1986. "Categorical shape and color coding by pigeons". *Journal of Experimental Psychology: Animal Behavior Processes* 12(2): 153–159.

Author's address

Angelo Cangelosi
Adaptive Behaviour and Cognition Research Group
School of Computing, Communication and Electronics
University of Plymouth
Drake Circus, Plymouth PL4 8AA
United Kingdom

acangelosi@plymouth.ac.uk
http://www.tech.plym.ac.uk/soc/staff/angelo

About the author

Angelo Cangelosi is Reader in Artificial Intelligence and Cognition at the University of Plymouth. His research interests include the modeling of language evolution in cognitive agents and robots, experimental and computational study of spatial language, the symbol grounding problem, and the combination of neural nets and genetic algorithms for the evolution of neural architectures. In Plymouth he leads the "Adaptive Behaviour and Cognition" research group.

Collaborative tagging as distributed cognition*

Luc Steels
Vrije Universiteit Brussel and Sony CSL Paris

The paper discusses recent developments in web technologies based on collaborative tagging. This approach is seen as a tremendously powerful way to coordinate the ontologies and views of a large number of individuals, thus constituting the most successful tool for distributed cognition so far.

Keywords: collaborative tagging, semiotic dynamics, distributed cognition

1. What is collaborative tagging

Without doubt, the most important recent development in the technological support for distributed cognition is the sudden explosion of collaborative tagging sites on the World Wide Web. The basic idea is straightforward. Users can upload or select certain materials (pictures, music files, texts, websites) and associate tags with these materials. The tags can be freely chosen and are similar to keywords. The user can then browse through tags. He can click on a tag and get materials that have been tagged the same way, see 'tag clouds' that graphically represent the popularity of tags, and see co-occurrence relations between tags and thus jump from one tag to another one (Figure 1).

Tagging sites began to appear in 2004, pioneered by the site delicio.us (for tagging websites), and very soon thereafter the site Flickr (for pictures). Collaborative tagging literally exploded in 2005 and is rapidly becoming a standard feature of websites. For example, Flickr went over the course of a year and a half from zero to close to one million users who uploaded and tagged about 20 million pictures. The past year growth rate has been 30% per month (Figure 2). Many additional sites have come on-line covering almost any domain for which there are websites today, including the exchange of scientific information (as the sites Cite.U.like or Connotea, managed by the Nature publishing

Pragmatics & Cognition 14:2 (2006), 287–292.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

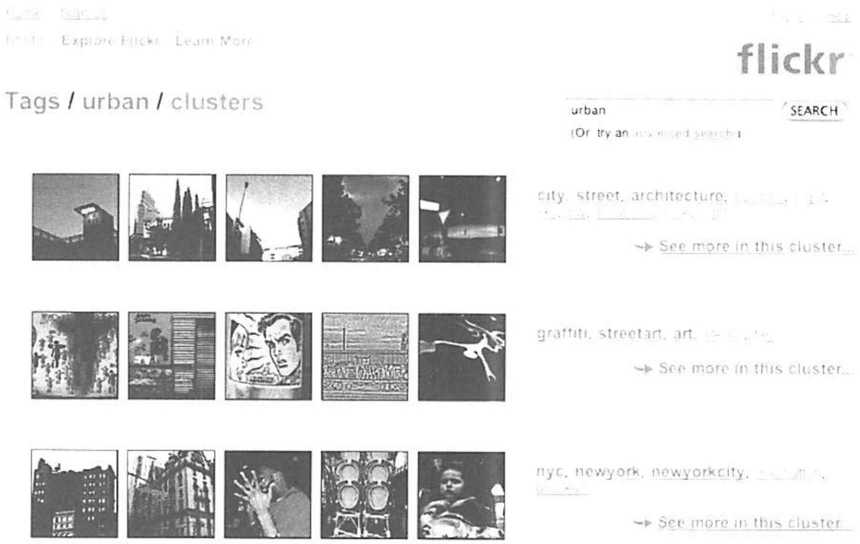

Explore Flickr
Learn More
flickr
Tags / urban / clusters
urban
SEARCH
(Or try an
city, street, architecture,
→ See more in this cluster...
graffiti, streetart, art,
→ See more in this cluster...
nyc, newyork, newyorkcity,
→ See more in this cluster...

Figure 1. Interface to the collaborative tagging site Flickr by which users can upload pictures and make them available to others. We see clusters of tags co-occurring with the tag 'urban'. The user can browse through pictures associated with each cluster or with specific tags.

group). Clearly collaborative tagging resonates very strongly with human user communities. How can we explain this remarkable development and why is it so interesting and relevant to research on distributed cognition?

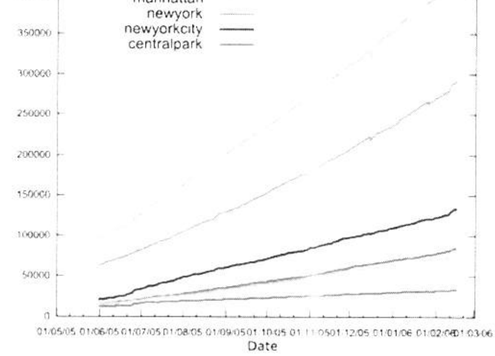


Figure 2. Phenomenal growth rate of pictures tagged with specific tags on the site Flickr for a period of 9 months. The tag 'nyc' went from 100,000 to 450,000 pictures. Notice also how the distribution between tags remains the same despite the scale up.

From a computer science perspective, tags are an alternative to the current URL hyperlinks used on the web. A hyperlink introduces a direct link to another site, and a user can simply click on a piece of text or on a picture and be transported to another site. But the designer of a site must know exactly the location of the link's target (expressed in terms of its URL, the 'Universal Resource Location' address) and the hyperlink itself must be in principle hard-coded in the text itself. This makes such links very precise, which is an advantage. But that can also be a disadvantage because it assumes more knowledge than may be available at the time a site is built and it is vulnerable to disappearing targets. Tagging is much more flexible because the relation between a tag and the address of the resource to which it may point is centrally managed and so new materials can become associated with tags in a distributed fashion. Users can introduce tags without knowing whether and how that tag has been used by others and a tag can attract new links even after it has been introduced by a particular user.

The second advantage of tags is that they introduce a kind of taxonomy of keywords or classifiers, which aid users in browsing by providing meta-data about the contents of uploaded material. So they offer an alternative to search engines like Google which perform search based on information retrieval over the actual contents of a web page. Proponents of the semantic web (Berners-Lee et. al. 2001) have been arguing for years that a new level of meta-data is necessary to carry the web further, but they succumbed to the temptation of handing the design of ontologies to centralised committees and of using logic with its insistence on total formal semantics and definition. Although there is now a sophisticated technology for designing such ontologies and formalisms for annotating websites, the idea of the semantic web has not taken off in practice, mainly because the effort and technological know how needed to design the ontologies is very high and reaching an agreement, even among experts, is notoriously difficult. Real world ontologies are simply too disorganised, fluid, and forever changing to cast into a logical harness. Of course, the semantic web technology is still useful and beneficial in rigid technical domains which are fairly static, like mathematics.

One would think that the freedom implied by open-ended tags would lead to complete chaos or a totally unusable system, but that is not the case. As recent studies that are tracking tag usage over longer periods of time have shown (Golder and Huberman, 2006; Cattuto et al., forthcoming), the distribution of tags for similar items stabilises rather quickly. Apparently a collective consensus arises and is then further maintained despite the absence of any global control. This can also be seen in Figure 2 where the tag 'nyc' remains dominant against competing tags like 'new york' or 'newyorkcity'. The statistical physicists that have carried out these investigations have argued that this emergent structure can be explained on the basis of statistical models, such as Polya's urn model or the Yule-Simon model, but this does not take into account the cognitive mechanisms and history that users bring to bear on collaborative tagging.

2. The relevance of tagging for distributed cognition

It seems more relevant to see tagging as an example of distributed cognition of the same sort that underlies natural language. For example, the 'tag clusters' reflect priming phenomena well established by psychologists. Just like a word triggers a barrage of other words for related concepts, a tag gets associated by simple co-occurrence with many other tags, and browsing through tag clusters makes sense for users because it resonates with their own associative semantic networks.

Several authors have recently argued that language should be viewed as a complex adaptive system in which a distributed group of agents collectively invent and align shared symbol systems. We now even have quite concrete models for this kind of semiotic dynamics, including for symbols that are grounded in reality through a sensori-motor apparatus (Steels 2003). These models should in principle help us to understand certain phenomena seen in collaborative tagging sites (Steels 2006). And indeed, they do. For example, these models show that a lateral inhibition dynamics induces synonymy breaking so that only a single word (or a few competing words) remain for the same sort of meaning. This phenomenon is also seen in tagging sites. Similarly the language game models suggest ways in which symbol grounding can be achieved through pattern recognition and signal processing and concrete experiments to ground tags in the real world are now ongoing that use the same ideas (Aurnhammer et al. 2006).

But collaborative tagging sites also teach us some new, profound lessons about how symbol systems emerge.

Work on the simulation of emergent communication systems so far focused on reference. Agents are motivated to look for and express properties of objects that are discriminating with respect to other objects. Tags are clearly not doing that. They appear to act like future aids for navigation in large information sources and the tagger does not know yet when and how a particular tag will be useful in the future. So we will have to invent other kinds of games in which navigation is the primary goal instead of reference. Second, tags are clearly not always groundable. When the tag 'nyc' is typed into Flickr, most of the images that are associated with this tag do not appear to have anything to do with New York City. They show for example a dog which has been photographed in an apartment of a friend in New York City. Similarly most pictures tagged 'red' do not show the color red at all. Tags therefore act as something in between symbols that stand for something in the world, and can therefore be grounded through perception and experience, and symbols that are purely used as labels or signposts to aid in navigation.

Today we are still lacking good models of how humans introduce symbols to organise their information spaces, and particularly how their personal choices are influenced by the choices of others. We also lack good models how symbols that were perhaps initially grounded become used as pure detached symbols and take on new meanings. Collaborative tagging is therefore a new rich source of empirical data and an enormous challenge for scientific models of distributed cognition.

Note

* This research was conducted at the Sony Computer Science laboratory in Paris, partly sponsored by the European FET project Tagora. I am indebted to Peter Hanappe for many discussions on tagging and for collecting the data shown in Figure 2.

References

Aurnhammer, M., Hanappe, P., and Steels, L. 2006. "Integrating collaborative tagging and emergent semantics for Image Retrieval". In *Proceedings of the Collaborative Tagging Workshop*. Web Conference, Edinburgh.

Berners-Lee, T., Hendler, J., and Lassila, O. 2001. "The Semantic Web". *Scientific American* 284(5): 34–43.

Cattuto, C., Loreto, V., and Pietronero, L. 2006. "Semiotic dynamics and collaborative tagging". *Nature* (submitted).

Golder S. and Huberman, B.A. 2006. "The structure of collaborative tagging systems". *Journal of Information Science* 32(2): 198–208.

Steels, L. 2003. "Evolving grounded communication for robots". *Trends in Cognitive Science* 7(7): 308–312.

Steels, L. 2006. "Semiotic dynamics for embodied agents". *IEEE Intelligent Systems* 21(3): 32–39.

Author's address

Luc Steels
6 Rue Amyot
75005 Paris
France

steels@arti.vub.ac.be
homepage: arti.vub.ac.be/~steels/

About the author

Luc Steels is professor in AI at the Free University of Brussels (VUB) and director of the Sony Computer Science Laboratory in Paris. He has contributed to many areas of Artificial Intelligence research ranging from knowledge-based systems to robotics. His most recent research topic is concerned with models for the origins and evolution of language using computer simulations and robotic experiments.

Thinking in groups*

Todd M. Gureckis and Robert L. Goldstone
Indiana University

Is cognition an exclusive property of the individual or can groups have a mind of their own? We explore this question from the perspective of complex adaptive systems. One of the principal insights from this line of work is that rules that govern behavior at one level of analysis (the individual) can cause qualitatively different behavior at higher levels (the group). We review a number of behavioral studies from our lab that demonstrate how groups of people interacting in real-time can self-organize into adaptive, problem-solving group structures. A number of principles are derived concerning the critical features of such "distributed" information processing systems. We suggest that while cognitive science has traditionally focused on the individual, cognitive processes may manifest at many levels including the emergent group-level behavior that results from the interaction of multiple agents and their environment.

Keywords: distributed cognition, emergence, complex adaptive systems, agent-based modeling, group problem solving, human foraging, social networks

1. Introduction

An implicit assumption made in cognitive science is that the individual is the critical unit of cognition. In recent years, the Distributed Cognition (DC) movement in psychology has arisen as a challenge to this assumption. Its adherents view the process of thinking as extending beyond the individual and instead as "distributed", either across the members of a group or in concert with objects and tools in the environment (Hutchins 1995a, 1995b). The DC perspective extends the boundaries of traditional cognitive science by including the larger network of ecological relations supporting the individual cognizer. However, one might reasonably question if such aggregate systems can themselves be productively viewed as cognitive, or if cognition is better understood from the perspective of the individual people comprising the system (cf. Harnad 2005).

Pragmatics & Cognition 14:2 (2006), 293–311.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

For example, when we say a group has "learned" something perhaps all we really mean is that some portion of the individuals in the group learned it.

There is an inherent tension between describing human groups in terms of the behavior of the individual versus the behavior of the aggregate. Focusing on the individual matches our personal experience. On a routine trip to the grocery store, we might be aware of the latest trend in the popularity of a breakfast cereal or the traffic on the way home, but we often have little sense of how our own goals or behavior might actually contribute to those phenomena. However, in even the most mundane daily activities, we become entwined in a network of aggregate social processes. Our cars contribute to the cyclic patterns of crowding and congestion on the highway (Helbing 2001) while our purchases interact with the consumption patterns of those around us to create trends and fashions (Gladwell 2000). These and other group-level phenomena such as rumors, the use of a standard currency, and the World Wide Web highlight the fact the behavior can be structured at levels beyond the individual.

Understanding the nature and cause of these types of group-level phenomena has been significantly advanced by the study of complex adaptive systems (Holland 1975; Resnick 1994; Kaufman 1996). One of the principal insights from this line of work is that rules that govern behavior at one level of analysis can cause qualitatively different behavior at higher levels. Thus, individual and group-level descriptions need not be in conflict. Armed with agent-based models, social scientists have developed detailed theories that explore how the low-level rules instantiated by interacting agents can account for organized patterns of behavior in the aggregate (Goldstone and Janssen 2005).

One of the most surprising results from this type of work is how the behavior of the group and the individuals in the group may diverge. For example, in the classic work of Thomas Shelling (1971), groups of agents who each have only a moderate bias towards living near members of their own "race" can give rise to extreme patterns of large-scale racial seggregation. Even though none of the agents wants to live in a completely segregated world, this is exactly the outcome of their individual actions and the cascade of reactions to those actions. Similarly, when a group must share a limited resource, egocentric individuals can cause all members of the group to be worse off than if they had acted in a more cooperative way (Axelrod 1984). These examples illustrate how attempts to understand individual behavior by observing the group can be misleading, as is trying to infer what the group will do based on the preferences of the individuals alone.

In our lab, we are interested in studying the collective behaviors that emerge when a large number of agents (both human and artificial) interact in a common world consisting largely of other agents. What are the critical properties of systems that exhibit interesting patterns of aggregate behavior? One critical feature seems to be that they are distributed, or spread across a set of multiple, interacting component units. In this sense, we consider the adaptive group-level behavior we observe in our work, and in the world at large, as a form of Distributed Cognition. Rather than tie cognition to first-person feeling as Harnad (2005) does, we (collectively, even more than we do individually) think that cognition is more productively construed in terms of adaptive problem solving. When a cockpit solves a navigation problem (Hutchins 1995b) or a university department successfully redefines its mission to fit changing times, we see these collective behaviors as evidence for adaptive information processing beyond the level of the individual human.

In what follows, we attempt to draw some lessons about the conditions under which the behavior of individual agents self-organize into adaptive group structures. First, we define what we see as some of the critical properties these systems. Then, we review a number of examples from our own lab where groups of people behave in adaptive ways and where the information processing achieved at the group level is qualitatively different from that of individuals in the group.

2. What makes something "distributed"?

Throughout science and engineering, there has been an intense interest in distributed systems. Biological evolution, nervous systems, the Internet, ant colonies, parallel computing systems, and traffic jams are all examples of distributed systems in action, and despite their considerable diversity there are a number of features they all share (Resnick 1994).

Perhaps most importantly, they are composed of identifiable units that can be used for different purposes and whose operation can be described independent of any particular context. Typically the operation of each unit depends on locally observable conditions rather than global information. The nodes in a computing cluster are all interchangeable, discrete systems which perform a number of functions independent of their inclusion in the network. Similarly, individual neurons in the brain have self-contained computational properties which depend largely on their local pattern of connectivity (Rumelhart and McClelland 1987). In human distributed systems, each person has their own set of perceptions, goals, and desires which can operate independently of the group. People are fairly independent because they can continue to function when isolated in the wilderness. To the extent that a person's functioning depends on their computer, desk, books, car, colleagues, and staff, their status as an independent information processing unit diminishes.

An implication of this characterization of units as structures independent of their context is that "unit-hood" is inherently graded. The status of a set of elements as a unit depends on two factors: the density of connections and dependencies among the set of elements, and the density of connections and dependencies between the elements within the set and elements outside of the set. The first factor decreases the unit-hood of the set of elements while the second factor increases its unit-hood. For example, a leading theory for the evolutionary origin of mitochondria and chloroplasts is that they were originally independent bacteria that became incorporated into the cytoplasm of cells, and once incorporated, conferred advantages for the cell because they allowed cellular respiration (mitochondria) and photosynthesis (chloroplasts) for energy production (Margulis 1970). We are less likely to view mitochondria as the individual units they once were because of their strong dependencies with other internal cell elements.

A second defining property of distributed systems is that the units are loosely coupled and can thus influence one another. This influence can take many forms. In computer systems this is the function of the network and communication protocol, while in human groups language, rhetoric, and social norms mediate interaction. Similarly, ants leave behind pheromone trails that other ants follow in order to find good foraging spots. However, other types of between-unit communication are more or less indirect. If we find out that a particular concert is sold out, a type of information has been transmitted from the individuals who bought the tickets to the individual wanting a ticket which may convey something about the perceived quality of the music.

A final property of distributed systems is that the pattern of connectivity between units is dynamic. If the elements of any system always belonged together, we would simply speak of them as one unit. Thus, in neural distributed system, synapse efficiencies, learning weights, or neurogenesis modulate new connection between units. In human groups, dynamic patterns of connectivity can be instantiated by changes to spatial or temporal proximity, or the pattern and strength of social relationships. In many cases, the pattern of connectivity can change considerably through time as individual units shift in and out of different groups.

The combination of these three properties is what makes distributed systems so interesting. For example, because of the loose coupling between units, the behavior of any single unit is contingent on the behavior of others.

The behavior of the group becomes the environment in which any single unit resides, thereby creating feedback loops of reactivity (i.e., because all the units around me do X, I will do Y which causes some of those units to do Z). In addition, due to the loose coupling and dynamic connectivity, distributed systems can implement a type of re-configurality. As units shift their allegiance between different groups, they allow cross fertilization. Units that have been influenced by their previous associations in turn bring these influences to bear on their subsequent interactions.

Distributed systems, then, are interesting because they exist at the *cusp* of unit-hood. Before the bacterium has been incorporated into the cell at all, it is simply an independent environmental influence on the cell. Once the mitochondrium loses its ability to make its own living in the world, it is no longer a unit by itself, but rather part of the eukaryotic cell unit. In between being a free-agent bacteria and a mitochondrial cog in the cellular wheel, the "bactondrial" is both independent and dependent on the cell. This status, we argue, is particularly important when it comes to cognitive systems. Computational complexity, in terms of being able to transmit information, is at its greatest for systems made of partially dependent elements. Sporns et al. (2004) have quantified the "information integration" of a system in terms of its total amount of mutual information. On the one hand, if a system's elements are completely independent, then information cannot be transmitted from one part of the system to another. On the other hand, if a system's elements are too tightly connected, then they all end up possessing the same information and communication is useless. Human nervous systems have apparently evolved so as to maximize the usefulness of neural communication (Sporns 2002). Similarly, we would argue that distributed systems incorporating people also adapt so as to create information-amplifying systems. Useful human collectives are those that promote robust information transmission across people yet avoid having everybody know the same things. Collectives that do this will maximize their computational capability.

Organization in such systems can be viewed from at least two levels: the operation of the collective, and the operation of the individual. These two levels need not be in conflict, although they are often not obviously related. By better understanding the operation of the individuals, one may derive considerable insight into the behaviors of the aggregate. Likewise, understanding the collective action of the group can give new insight into the behavior of the individuals.

3. Some case studies

Of course, simply being "distributed" in the sense just described in no way guarantees that any particular system will behave in any particular way. In some systems, the action of the group is quite simply the straightforward action of the individuals. However, in other systems new higher-level, emergent properties develop. In order to better understand what features encourage this process, our lab has developed an interactive computer platform which allows a large number of individuals (and agents) to interact in real-time.[1] Through this system we have been able to systematically study the effects of group behavior as it is driven by the environment, goals, perceptions, and motivations of individuals in the group. In what follows, we review a number of case studies from this work demonstrating the conditions under which adaptive group behavior can arise.

3.1 Group path formation

In one set of studies, we explored how individuals in a group are influenced by the choices and behaviors of their predecessors in the development of path or trail system (Goldstone, Jones, and Roberts, in press). Often times, there are advantages to choosing options which other's have already established as popular or useful. Early trail blazers through a jungle use machetes to make slow progress in building paths, progress that is capitalized on and extended by later trekkers, who may then widen the trail, then later put stones down, then gravel, then asphalt.

In our experiments, a small group of participants (6–12 people) interacted in real-time through a collaborative computer system. Each individual's task was to control a green triangle on the screen using the arrow keys on a keyboard and to direct it to a number of different goal positions on the screen. Navigation between these various goal states was complicated by a cost function which deducted points depending on where the participants moved. The number of points deducted on each step decreased as a function of the number of times that particular location had been stepped on by all participants of the game. Since the goal for each participant was to navigate to each location while losing the least number of points, they had to balance the advantage of short beeline paths between two destinations against the advantage given by traveling in regions that were popular with the group (i.e. established trails). The relative cost of each position in the virtual world was made visible to each participant by the color of each cell (higher cost cells were darker in the display, see Figure 1).

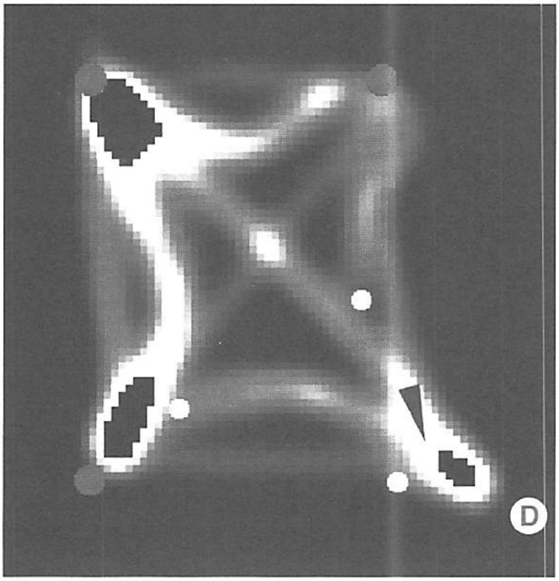


Figure 1. In this figure, the participants' own location is shown is shown by the dark grey triangle. The other agents are shown as small white circles, and the destinations are shown as the dark grey circles. The target destination for the participant is shown as a bright white circle labeled with the letter D, and the ease-of-travel of each location on the world is shown by its moment-by-moment brightness.

Over time, the participants combine their efforts to establish "trails" in this system just like the spontaneous paths that cut across the grassy areas on a university campus. One question of interest was if agents following the emergent paths of the group end up traveling less overall distance to reach to their respective destination (i.e. does the group find optimal path structures for connecting the destinations?). Fortunately, there is a relatively straightforward characterization of the shortest system of paths that connect any number of destinations called the Minimal Steiner Tree (MST). An example of MST solutions for two configurations of destination points is shown in Figure 2. The MST solution is the one that minimizes the total amount of "path" which is required to connect all the destinations in the configuration. In many configurations this is accomplished by adding some number of new, intermediate destinations to the configuration called Steiner Points.

The results showed that while groups did not necessarily find the best MST solution for any particular configuration, there were deviations from beeline paths towards what would be a MST solution. These pro-MST deviations were particularly common in places where the structure of the environment

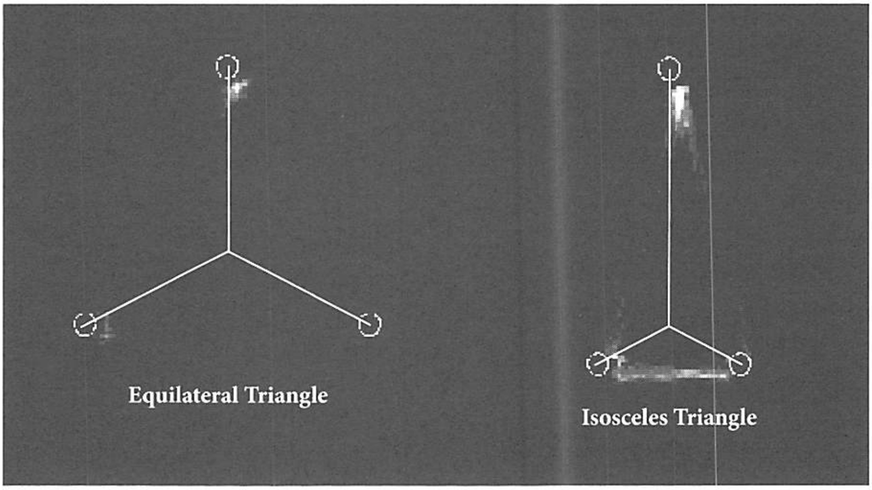


Figure 2. The cumulative steps taken on each cell for the Equilateral and Isosceles triangle configurations. The brightness of a location is proportional to the number of times that it was stepped on by all participants. The destinations are indicated by white circles. The shortest path system connecting the destinations is shown by the three white lines, and is known as the Minimal Steiner Tree.

encouraged path reuse such as in the isosceles triangle show in Figure 2. In comparison, the equilateral triangle arrangement of destinations showed relatively little deviation from the set of beeline paths. For the equilateral triangle, there is no incentive for a "pioneer" to move straight down from the top-most destination — it is dark and costly there. However, for the isosceles triangle arrangement, a participant moving from the top-most destination to the bottom-right destination might have an incentive to use the left pathway coming out of the top-most destination if, by chance, this path is somewhat bright because it has been recently traveled. Then, the participant will cut across at some point to the right-most city, thereby stepping in the middle territory. Once they have stepped in the middle region, it becomes somewhat more tempting for further travelers, who exploit and extend still further this middle, vertical trail (see Figure 2). The traveled path system for the isosceles triangle "zips up" over time. Participants originally start traveling by the beeline paths, but gradually step in the center regions, which causes still more steps in the center regions. Interestingly, the largest savings of distance for the Minimal Steiner Tree (shown as white lines in Figure 2) compared to the beeline paths is for the equilateral rather than isosceles triangle. Empirically, this is exactly the configuration that produces the fewest steps away from the beeline paths. Thus, having a strong group-level advantage for a path system is no guarantee that the group will find that path system.

To understand why subjects made the paths they did it is important to keep in mind that the environment in which the paths were formed was not simply the surface on which they "walked", but the interaction of the surface and the behavior others in the task. As individuals sometimes at random venture into new regions of the space, it becomes easier for subsequent agents to travel there. Only through a collective process of exploration and exploitation do agents arrive at a particular path solution. Our results showed how very slight modifications to the environment could encourage the group as a whole towards more optimal solutions. In particular, the structure of the isosceles triangles encouraged participants to move away from beeline paths by first fostering small steps away from the beeline paths, and then having subsequent travelers extend still further these deviations.

These results suggest that subtle changes to the structure of the environment and the incentives each agent responds to can have considerable influence on the group as a whole. Instead of using laws and strict barriers in order to control crowds, these less authoritarian methods of crowd control can be just as effective. For example, each year over 2.5 million people make their way across the Jamarat Bridge during their pilgrimage to Mecca. In 1990, as many as 1400 people were crushed to death by the massive crowd. Hughes (2003) applied methods from fluid dynamics to model the flow of pilgrims across the bridge and his models found that a series of barriers which helped to direct the density and speed of the crowd could effectively prevent congestion at dangerous spots.

3.2 Propagation of innovations

Related to the issue of how individuals build upon the work of others in constructing spatial paths is the propagation of innovations in more abstract domains. In solving problems, individuals often have to balance the cost of acquiring their own information against simply copying the successful behaviors of someone else (Bandura 1965). In order to understand the factors that influence the propagation of innovative solutions within a group, members of our lab have explored how groups pool and exchange information in order to solve an abstract problem (Mason, Jones, and Goldstone, in press).

Participants attempted to find the solution to a problem involving a simple, continuous number space. In each of 15 successive rounds, subjects guessed a number between 0 and 100 and were shown a numerical score as feedback that indicated how good their guess was according to a hidden evaluation function (see Figure 3). Each participant's goal was to maximize their own cumulative

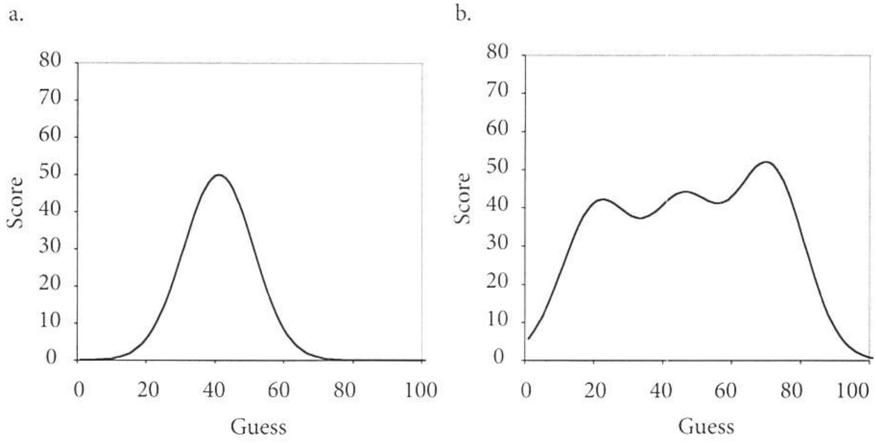
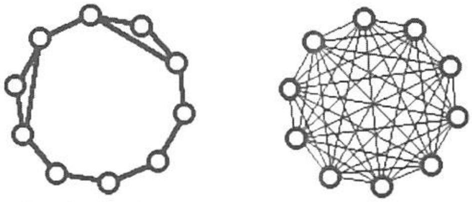


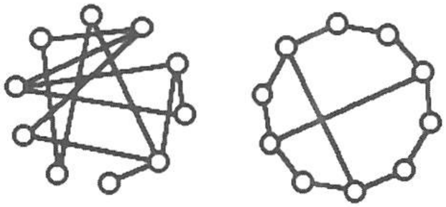
Figure 3. Examples of the (a) unimodal and (b) trimodal fitness functions which convert guessed numbers to scores.

number of points earned across the 15 rounds. In addition to receiving feedback about their own guesses, participants were also shown the guesses and scores of people who they were directly connected with. Thus on each round participants could explore near their own previous guesses, or switch to the successful guesses of people with whom they were connected.

The experiment manipulated two factors: the pattern of connections between individuals and the structure of the problem the individuals were to



Regular Lattice Fully Connected



Random Small World

Figure 4. Examples of the different network structures for groups of 10 participants. Circles represent participants and lines indicate communication channels.

solve. There were four different network structures studied: full, lattice, random, and small world (see Figure 4). In the full network, every person was connected to every other person in the game. Thus, at the end of each round of the experiment, each participant had knowledge about the quality of the solution given by every other person. In the lattice condition, people were connected to only a small number of immediate neighbors. This network causes a large amount of local clustering in that a good solution found by one participant in one part of the graph must travel a long distance in order to reach people in other parts of the graph. In the random graph condition, people were connected randomly to other individuals. Random graphs have less clustering, and lower average path length between individuals, which speeds the flow of information compared to the lattice. The final condition was a small world network, which falls somewhere between the extremes of the highly structured lattice and the random graph (Watts and Strogatz 1998). Small world network combines high degrees of clustering with long range "short-cuts" which lower the average shortest-path length. The problem spaces, shown in Figure 3, were either unimodal (having one "best" solution) or trimodal (having two local maxima in addition to the globally best solution).

The principal factor of interest was how fast the group converged on the best solution, and how complete the convergence was (in terms of the percentage of individuals within some close proximity to the optimal solution). In the unimodal problem space, the networks with the shortest path average path length found the solution most quickly (i.e., the small world, random, and full networks). The full network converged upon the best solution most quickly of all. This matches the intuition that with such a simple problem, the group that can broadcast the solutions most effectively would perform best.

Performance in the trimodal problem space found a different pattern of results. In these sessions, the best solution was found faster, and by more individuals in the small-world condition compared to the full, random, or lattice networks. The structure of the trimodal problem space and the structure of small world networks align to help the group solve the problem. The large amount of clustering in the small world network allows for independent searching in different parts of the network enhancing the exploration of the problem space. On the other hand, the relatively short path length in the small-world networks allows the effective distribution of good solutions once they are found. In effect, the small-world network balances the competing demands of exploration versus exploitation.

Perhaps counter-intuitively, this result suggests that less information can actually be more when a group collectively searches a problem space. One hypothesis might have been that the full network would always have an advantage because it allows the most collaboration and exchange. However, the short path length and strong clustering of the small world network allowed for more regionally-based specialization within the overall group, leading to a more robust search of the entire problem by the group.

3.3 Human foraging behavior

Effective search is a critical part of human cognition. In the study just reviewed, search transpired in an abstract problem space. How do humans search for and exploit more concrete resources, and are there parallels between searches of spatial and abstract problem spaces (Hills 2006)? A final set of case studies conducted by members of our lab explored how humans search for and allocate themselves to resources (Goldstone and Ashpole 2004; Goldstone, Ashpole, and Roberts 2005).

As in the group path experiments, participants in these studies controlled a small character on the screen, however this time their goal was to collect resources which appeared at two different sites on the screen. Participants gained points for each piece of "food" they collected by stepping on it before any other person. The two resource pools had different experimenter-manipulated payoffs (50–50, 65–35, and 80–20). In addition, various conditions manipulated the information available to participants at any time. Participants could either see the location of other participants on the screen, the location of available food, both of these, or neither of these.

The question of interest was how the agents in the group would distribute themselves between these various resources. In biology there are models for how organism distribute themselves amongst a set of resources called the Ideal Free Distribution model (IFD) (Fretwell and Lucas 1972). Basically, this model predicts matching of agents to resources — if twice as much food is available at location A compared to location B, then the model predicts twice as many animals at location A compared to location B.

The results from these experiments found significant violations of the IFD model. In particular, in almost all conditions there was undermatching (i.e., if the resources were distributed 80/20, then participants distributed themselves 70/30). Undermatching implies that as a group, participants did not take advantage of the resources in an optimal way. The only exception to this was in the condition where resources were visible, but other agents were not. In this case, participants tended to overmatch (i.e. slightly more agents gravitated towards the more plentiful resource than predicted by the IFD model). Overall, the empirical results and associated computational modeling (Roberts and Goldstone 2005) indicate that people use knowledge of food density to move to lucrative regions, and use knowledge of the whereabouts of other foragers to distance themselves from crowds. However, we also found evidence for "bandwagon behavior" — when food was invisible but other participants were visible, then foragers used the appearance of other foragers as evidence that a region might be productive.

Perhaps the most interesting empirical result has been oscillations in the harvesting rate of the resources pools across time, particularly when participants had no knowledge of the location of either other foragers or of available food. Fourier analysis revealed strong fluctuations in the utilization of a particular pool in the range of 50 seconds per cycle. These waves of crowding were caused by correlated dynamics in the decision making of individual agents. Initially, an appealing resource patch would become overcrowded with foragers. This crowding would lead to relatively low payouts to the individuals in the crowd. This, in turn, led to an extensive migration out of the patch, making the patch, once again, attractively underused, thereby completing one cycle of population flow. These cycles were not found when other foragers were visible because the temptation to leave an over-harvested resource pool would be tempered by the realization that several others have already begun the migration to the less crowded pool. The irony of the waves of crowding in the conditions with invisible foragers is that participants are highly motivated to avoid traveling with the mob. This individual-level motivation for each forager to be in the minority is exactly what leads the group to travel as a mob!

4. Lessons learned

There are a number of lesson learned from these experiments:

People are a large part of people's environments

Consistent with the DC perspective, the behavior of individuals is influenced and supported by an environment composed of other people. In each of our experiments, the behavior of most individuals followed that of the group, despite the fact that people were free to act however they wanted. Rather than build their own trails, or seek their own problem solutions people naturally take advantage of the behavior of others when devising their own actions.

The interaction between individuals in our experiments was not always direct. The biologist Pierre-Paul Grassé (1959) developed the notion of "stigmergy": how organisms often communicate indirectly by modifying their local environment. This type of communication was prevalent in our studies where the tracks left by early travelers invited later travelers to follow. Similarly, in the forager experiments, the waves of oscillation between foraging spots occurred when individuals, by virtue of over-harvesting their local environment, changed its intrinsic value to the group.

Divide and conquer: Exploration and exploitation in groups

In many systems, effective management depends on leaders who delegate necessary tasks to their subordinates. However, our experiments show conditions in which a group may spontaneously organize into an effective problem solving structure without centralized control. In searching a hidden problem space with local maxima, groups connected by small world networks naturally took advantage of their distributed nature to both explore and exploit the problem domain. Given that the tension between exploration and exploitation in learning systems is a common challenge facing intelligent systems (Sutton and Barto 1998), it is interesting that our groups were able to arrive at solutions which balanced the two.

More information isn't always better

The amount of information that individual agents have access to clearly influences the efficiency of the group. When both resources and other agents were invisible, the groups in the foraging experiments made less optimal use of the food distributions than when more information was provided. However, more information is not always better. In the group problem solving experiments, the small world network provided more robust search abilities than did the full networks. Here less information encouraged independence between individual agents that greatly enhanced group performance in a complex problem with multiple local maxima.

Influencing groups by bottom-up pressures rather than top-down rules

Finally, our experiments suggest new ways for control the behavior of groups. The most common method of crowd control is through direct orders or laws. If we wish to direct pedestrian traffic, for example, we may institute rules or physical barriers that prohibit certain movements. Rather than enforcing restrictive laws or boundaries, minor changes to the incentives of even a few individuals can cause radical changes in the group. For example, in our path formation experiments, even without instituting physical or abstract barriers, it may be possible to indirectly control collective behavior with substantial efficacy. The

"Active Walker" model in biophysics (Helbing, Keltsch, and Molnár 1997) does a good job of explaining and predicting our empirical results. In this model, walkers move to destinations, and as they take steps, they affect their environment, facilitating travel for subsequent walkers. Walkers compromise between taking the shortest way to their destination and using existing, strong trails. Two critical parameters of this model are how quickly the influence of a step dissipates, and how visible the strengths of trails are to walkers. As influence dissipation becomes more rapid, and as path visibility increases, the agents collectively form better approximations to the optimal Minimal Steiner Tree path systems. For situations where conserving the total amount of pathway is desirable (e.g. when vegetation must be cut down to create the paths), planners should explore ways of increasing path visibility, the efficacy of steps, or path decay. Given the empirical success of the Active Walker model with our groups of people, varying its parameters becomes a potentially useful way of not only predicting, but controlling the growth of spontaneous paths.

Collective behavior is potentially more controllable than isolated individual behavior because of the strong influences among the individuals' behavior. A small pressure can often be magnified by the positive feedback involved in individuals following other individuals (Dorigo et al. 2000). This has important implications and warnings for public policy. On one hand, the incentives given to individuals do have important impact on the behavior of the group. Thus, tax policies that encourage certain behaviors can lead to desired social changes. On the other hand, the relationship between the individual biases and group-level behavior is often not obvious. In the group path formation experiments, despite incentives to cooperatively lay down efficient trails, the groups tended not to find the optimal solution. Agent-based modeling provides a new avenue for exploring the implication of various government policies on the emergent behavior of collectives (Epstein and Axtell 1996).

5. Conclusions

Earlier we brought forward a number of features we believe characterize distributed systems. In particular, we argued that distributed systems that strike a critical balance between the dependence and independence of their units would exhibit unique information processing characteristics. This characterization is well exemplified by the reviewed findings. By manipulating the pattern of connectivity between individuals in the group problem solving experiment, we found that the most robust group search strategies came when people were connected in small-world networks, which balance strong local dependencies against global connectivity. Similarly, the most optimal group paths were formed when the structure of the environment encouraged path re-use (i.e., the isosceles configuration) and thus interactions between the path histories of separate individuals following their own goals. When there was little chance for overlap, such as in the equilateral configuration, the group tended to be less efficient and instead simply reflected the expected paths of non-interacting agents. By being only partially dependent, elements avoid becoming redundant. The dangers of redundancy are well illustrated by the tendency of people connected via a full network to converge on sub-optimal local maxima. The price of everybody being connected to everybody else directly is that a federation of explorers comes to behave as a single explorer.

Perhaps most importantly, the studies reviewed show several dissociations between individual- and group-level behavior. For example, the individual foragers in a group with other invisible foragers are motivated to be in as underpopulated a resource pool as possible. This motivation leads them, perversely enough, to travel in crowds. In the group search problem when individuals can see everybody else's scores they end up converging quickly on a single solution, rather than exploring the full range of solutions. The result is that individual attempts to maximize points earned lead to poor group performance. In these examples, individual actions have unanticipated and undesirable consequences for the group's behavior.

The lessons we found in our work have implications for how distributed systems should be designed, deployed, and controlled. Unlike traditional top-down systems, apparently innocuous design decisions may greatly affect global behavior in distributed systems. For example, maintaining a critical level of independence for individual units is important for teams, committees, and economies in order to most effectively search a complex problem space. The often non-intuitive relationship between individual motivations and group behavior highlight the importance in building computational models to test the impact of particular design decisions. As computational simulations become increasingly realistic, they will serve as increasingly useful test-beds for exploring the potential consequences of public policies that have complex, non-linear dynamics.

Let us return to the question of whether our experiments demonstrate genuine distributed *cognition*. We have no way of assessing Harnad's (2005) criterion that the group actually *feels* something, any more than a neuron can assess whether its human owner feels. However, our experiments demonstrate how the behavior of interacting groups of people may organize into an adaptive, problem solving structure. This creative, adaptive problem solving behavior is a key characteristic of cognitive processing, and so it is not unreasonable to consider such group behavior a form of distributed cognition. The solutions of our groups varied in quality. The collective travelers do not ever form Minimal Steiner Trees while the foragers actually do distribute themselves with considerable efficiency to resources. However, regardless of the quality of solutions, the important point is that the group-level behavior can be evaluated separately from the individual-level behavior. It has a life of its own.

The complex systems framework provides a powerful tool for understanding the behavior of a diverse set of natural systems, including cognition. The idea that systems composed of loosely coupled, but well defined units can self-organize into systematic behaviors at higher levels of analysis closely parallels the ethos of Distributed Cognition which seeks descriptions of cognition at both the level of the individual and the group. Understanding how these two levels of description mutually reinforce each other is an important step toward richer theories of many natural phenomena.

Notes

* This work was supported by NIH-NIMH training grant #:T32 MH019879-12 to T.M. Gureckis, and Department of Education, Institute of Education Sciences grant R305H050116, and NSF REC grant 0527920 to R.L. Goldstone. We would like to thank Michael Roberts, Andy Jones, Winter Mason, and Tom Wisdom for intellectual contributions.

1. Participate online at http://groups.psych.indiana.edu/

References

Axelrod, R. 1984. *Evolution of Cooperation*. New York: Basic Books.

Bandura A. 1965. "Behavioral modification through modeling procedures". In L. Krasner and L.P. Ulmann (eds.), *Research in Behavior Modification: New Development and Implications*. New York: Rinehart and Winston, 310–340.

Dorigo, M., Bonabeau, E., and Theraulaz, G. 2000. "Ant algorithms and stigmery". *Future Generation Computer Systems* 16: 851–871.

Epstein, J.M. and Axtell, R. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brookings Institution and The MIT Press.

Fretwell, S.D. and Lucas, H.J. 1970. "On territorial behavior and other factors influencing habitat distribution in birds: I. Theoretical development". *Acta Biotheoretica* 19: 16–36.

Gladwell, M. 2000. *The Tipping Point*. New York: Little, Brown, and Company.

Goldstone, R. and Ashpole, B. 2004. "Human foraging behavior in a virtual environment". *Psychonomic Bulletin & Review* 11: 508–514.

Goldstone, R.L., Ashpole, B.C., and Roberts, M.E. 2005. "Knowledge of resources and competitors in human foraging". *Psychonomic Bulletin & Review* 12: 81–87.

Goldstone, R. and Janssen, M. 2005. "Computational models of collective behavior". *Trends in Cognitive Sciences* 9(9): 424–429.

Goldstone, R., Jones, A., and Roberts, M. In press. "Group path formation". *IEEE Transactions on System, Man, and Cybernetics, Part A.*

Grassé, P.-P. 1959. "La reconstruction du nid et les coordinations interindividuelles. La théorie de la stigmergie". *Insectes Sociaux* 6: 41–84.

Harnad, S. (2005) "Distributed processes, distributed cognizers, and collaborative cognition". *Pragmatics & Cognition* 13(3): 501–514.

Helbing, D., Keltsch, J., and Molnár, P. 1997. "Modeling the evolution of human trail systems". *Nature* 388: 47–50.

Helbing, D. 2001. "Traffic and related self-driven many-particle systems". *Review of Modern Physics* 73: 1067–1141.

Hills, T. 2006. "Animal foraging and the evolution of goal-directed cognition". *Cognitive Science* 30: 3–41.

Holland, J. 1975. *Adaptation in Natural and Artificial Systems.* Cambridge, MA: The MIT Press.

Hughes, R. 2003. "The flow of human crowds". *Annual Review of Fluid Mechanics* 35: 169–182.

Hutchins, E. 1995a. *Cognition in the Wild.* Cambridge, MA: The MIT Press.

Hutchins, E. 1995b. "How a cockpit remembers its speed". *Cognitive Science* 19: 265–288.

Kaufman, S. 1996. *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity.* Oxford: Oxford University Press.

Margulis, L. 1970. *Origin of Eukaryotic Cells.* New Haven, CT: Yale University Press.

Mason, W.A., Jones, A., and Goldstone, R.L. In press. "Propagation of innovations in networked groups". *Journal of Experimental Psychology: General.*

Resnick, M. 1994. *Turtles, Termites and Traffic Jams: Exploration in Massively Parallel Microworlds.* Cambridge, MA: The MIT Press.

Roberts, M.E. and Goldstone, R.L. 2005. "Explaining resource undermatching with agent-based models". In *Proceedings of the Twenty-seventh Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum, http://www.cogsci.rpi.edu/CSJarchive/Proceedings/2005/index.htm.

Rumelhart, D. and McClelland, J. 1987. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Cambridge, MA: The MIT Press.

Schelling, T. 1971. "Dynamic models of segregation". *Journal of Mathematical Sociology* 1: 143–186.

Sporns, O., Chialvo, D., Kaiser, M., and Hilgetag, C.C. 2004. "Organization, development, and function of complex rain networks". *Trends in Cognitive Sciences* 8: 418–425.

Sporns, O. 2002. "Network analysis, complexity, and brain function". *Complexity* 8: 56–60.

Sutton, R.S. and Barto, A.G. 1998. *Reinforcement Learning: An Introduction.* Cambridge, MA: The MIT Press.

Watts, D.J. and Strogatz, S.H. 1998. "Collective dynamics of "small-world" networks". *Nature* 393: 440–442.

Authors' addresses

Todd Gureckis or Robert Goldstone
Indiana University
Department of Psychology
1101 E. 10th Street
Bloomington, IN 47405

tgurecki@indiana.edu
rgoldsto@indiana.edu

About the authors

Todd Gureckis is a postdoctoral fellow at Indiana University on the NIH-NIMH Cognitive Modeling Training Grant. He received his Ph.D. in psychology in 2005 and his B.S. in Electrical Engineering in 2001 both from the University of Texas at Austin.

Robert Goldstone is a Chancellor's Professor of Psychological and Brain Sciences and the Director of the Cognitive Science program at Indiana University. He was awarded a 2000 APA Distinguished Scientific Award for Early Career Contribution in Cognition and Learning, and a 2004 Troland research award from the National Academy of Sciences.

Distributed learning and mutual adaptation*

Daniel L. Schwartz and Taylor Martin
Stanford University / University of Texas at Austin

If distributed cognition is to become a general analytic frame, it needs to handle more aspects of cognition than just highly efficient problem solving. It should also handle learning. We identify four classes of distributed learning: induction, repurposing, symbiotic tuning, and mutual adaptation. The four classes of distributed learning fit into a two-dimensional space defined by the stability and adaptability of individuals and their environments. In all four classes of learning, people and their environments are highly interdependent during initial learning. At the same time, we present evidence indicating that certain types of interdependence in early learning, most notably mutual adaptation, can help prepare people to be less dependent on their immediate environment and more adaptive when they confront new environments. We also describe and test examples of learning technologies that implement mutual adaptation.

Keywords: distributed cognition, learning, education, early mathematics, instructional technology, small group interaction

1. Introduction

If distributed cognition is to become a general analytic frame, it needs to handle more aspects of cognition than just highly efficient problem solving. Thus far, distributed cognition has mainly described how environments and people are components of larger systems that have been designed to accomplish specific informational tasks. Another critical aspect of most cognitive systems, however, is that they learn. Does it make any sense to consider distributed learning where both individuals *and* their environments learn, or shall we restrict learning to a non-distributed analysis and relegate distributed cognition to narrow application?

Pragmatics & Cognition 14:2 (2006), 313–332.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

To make this a fair question, we should probably describe learning as functional adaptation, so it can apply to both humans and the environment. Given an adaptation definition of learning, a strong case for distributed learning will require instances of individual adaptations that are difficult to describe if we ignore adaptations to the environment (whether physical or social). Furthermore, to make the case, it should be possible to generate and test learning predictions that stem from a distributed framework. Finally, if the idea of distributed learning has practical value, then it needs to guide the design of environments to support learning. In the following, we try to provide a preliminary analysis, empirical evidence, and learning technologies to make the case for distributed learning and a fuller account of distributed cognition. Our basic point is simple. Most cognitive research has been silent about the signature capacity of humans for altering the structure of their social and physical environment. Nevertheless, the adaptation of the environment is a powerful catalyst for learning and self-adaptation, and instructional designers should capitalize on this capacity so as to improve learning experiences.

2. Distributed learning and education

Our interest in distributed learning grows from our interest in education. Cognitive research in education has faced a difficult problem. Educators need to prescribe environments that orchestrate specific learning processes. The demands of educational research may be contrasted with those of cognitive psychology and neuroscience, where the focus is usually on the internal mechanisms of thought. Here, the methods of research use a particular environment to trigger internal processes, but the environment is not part of the theory about how people think. Unfortunately, it is difficult to design learning experiences solely on a theory of internal process. It rarely works to say, "People will learn using mental process X in an environment that is like the one they used in Bob's study, all things being equal". All things are rarely equal. Environments differ across classrooms. They also change over time. This is one of the appeals of research on distributed cognition. It is working towards theories that encompass how individuals and environments interact.

A theory of distributed cognition relevant to education *must* consider learning. Most of the work in distributed cognition has focused on problem solving. Researchers examine how information processing and storage are distributed across people and contexts (for a review, see Pea 1993). This has been useful for exploring how people efficiently accomplish familiar tasks with the aid of smartly constructed environments. Improved computer interfaces, for example, can help people off-load the demands of recalling how to initiate a spreadsheet process. This type of research does not explain learning, or how these environments became distributed the way they are. Though learning and problem solving strongly influence one another, learning yields lasting change, while problem solving does not. Thus, we distinguish distributed *problem solving* from distributed *learning*.

3. Four types of distributed cognition for learning

To help clarify situations that involve distributed learning, we created Figure 1. The figure characterizes types of learning by the relative degree of stability or adaptability of individuals and their environments.

Individual	Adaptable	(1) Induction	(4) Mutual Adaptation
	Stable	(2) Symbiotic Tuning	(3) Repurposing
		Stable	Adaptable
		Physical and Social Environments	

Figure 1. Four Quadrants of Distributed Cognition (Adapted from Martin and Schwartz 2005)

In Quadrant 1 — Induction — people do not have stable, mature ideas, but they are operating in well-structured and stable environments. By induction, we do not mean a specific psychological process or learning algorithm. Rather, we mean it more broadly in the sense that a current in one wire can induce a current in another wire. Abacus masters, for example, have expansive digit spans and arithmetic abilities because they have induced a mental model that can simulate the structure of an abacus (Hatano and Osawa 1983). In Quadrant 1 people induce the regularities of a stable environment. In physical environments, the learning interaction often involves probing an environment to infer its structure. In social environments like a classroom, the induction might involve trying to infer the meaning behind a teacher's words. Glenberg et al. (2004) provides an excellent example of Quadrant 1 learning. Early readers learned to create mental models of text passages by moving characters in a physical scene.

In Quadrant 1, learning involves changes to the individual. The environment does not change its basic structure, though it can be moved around.

In Quadrant 2 — Symbiotic Tuning — people operate in stable and often specialized environments *and* they have stable ideas. For example, highly trained pilots operate in the well-designed environment of a cockpit. They form a system that succeeds at achieving complicated feats such as remembering airspeeds (Hutchins 1995). Activity in this quadrant has been the mainstay of the distributed cognition literature, where the question is how people off-load intellectual work to the environment (Norman 1988; Zhang and Norman 1994). In the symbiotic quadrant, learning is characterized by an increase in the efficiency of the system and the interdependence of its components. Because the components become increasingly interdependent, the learning is often bound to the context. Pilots cannot transport their specialized reliance on the cockpit to another environment, nor can a highly specialized cockpit transport to another setting; hence, the label "symbiotic tuning". Adaptations to the individual and environment depend on one another to manifest themselves.

In Quadrant 3 — Repurposing — people have stable ideas, but the environment does not have an ideal form. If people's ideas are mature enough, they can repurpose the environment to serve their functions (Kirsh 1996). For example, people might repurpose a knife to turn a screw, or invent a screw in the first place. Repurposing involves adapting the environment to implement one's ideas. Once the environment stabilizes, individuals can move into a more symbiotic mode and improve the efficiency of the system.

In Quadrant 4 — Mutual Adaptation — people do not have stable ideas nor do they operate in a fixed environment. While ideas and environments do not change constantly, people often need to adapt their environments and ideas. A home gardener may need to switch from roses to shade-loving plants. In the process of adapting the garden, the gardener will learn more about shade-loving plants. Mutual adaptation involves the co-evolution of the individual and the environment.

4. A hypothesis about mutual adaptation

In the following, we provide several examples of distributed learning to indicate why we think it is relevant to education. In the first examples, we consider physically distributed learning. We consider cases where children manipulate physical objects to help them learn math concepts. In the second examples, we consider socially distributed learning. We consider cases where learners adapt with other people.

For both physically and socially distributed learning, we have a special interest in Quadrant 4: Mutual adaptation. One reason is that we have found examples of learning that would be incompletely characterized if we only focused on individual adaptations (described below). A second reason is that we have found that opportunities for mutual adaptation provide important benefits to students compared to the more canonical situation where the environment is fixed and only the student can change. In particular, we will show that mutual adaptation prepares learners to transfer their adaptations to new environments. We end each section below with an example of how our perspective on mutual adaptation leads to implications for learning technologies.

4.1 Physically distributed learning

For our discussion of physically distributed learning, we consider the case of hands-on manipulation in mathematics learning. This is an interesting test case for two reasons. First, students often use hands-on materials (manipulatives) in early mathematics, but there is quite a bit of academic and political confusion about how to conceptualize their role in learning. Second, it is not a trivial question how children can work with concrete materials to learn abstract ideas, as opposed to just learn more concrete operations.

Our particular case involves children learning about ratio. Ratio is a common bottleneck in arithmetic learning. When children meet formal treatments of ratio around 8-years old, they already have highly efficient whole-number schemas for counting and other basic number operations. Children often interpret ratio problems in terms of these schemas. When asked to solve symbolic problems, like 1/2 + 1/4, children often answer 2/6 (Carpenter, Corbitt, Kepner, Lindquist, and Reys 1980; Kerslake 1986). Given pictures or hands-on materials, they make similar mistakes. For example, if children are shown eight pieces and asked to indicate one-fourth, they often indicate one piece, four pieces, or both.

For ratio problems, children need to develop new interpretations of quantity. For example, to solve 1/4 of 8 pieces, they need to reinterpret the pieces as groups, so they can map the "4" of the problem into four equal groups. They can then map the "1" of the fraction into one group, and count the number of items within the group. This level of reinterpretation is difficult to do mentally, especially if one does not know that such an interpretation exists (Chambers and Reisberg 1985). This is where physical objects may be useful; children can easily and quickly rearrange manipulatives, and this reorganization of their environment could lead to adaptations in their interpretations.

Most applications of manipulatives fall into Quadrant 1 (Induction) where ideas are adaptable and the environment is taken as stable. The cognitive assumption is that manipulatives provide children direct access to quantitative structures. Once induced, children will be able to map the structures from the physical domain into the symbolic domain (e.g., Piaget 1953; Dienes 1973). Instructionally, the idea is that manipulatives provide a concrete and easily accessible analogy for symbolic procedures (e.g., Hall 1998). Base-10 blocks are a good example. A ten-rod equals 10 unit-blocks, a hundred-plate equals 10 tenrods and 100 unit-blocks, and so on. The base-10 blocks have a stable structure that mirrors the base-10 notational system. Using base-10 blocks, children are supposed to learn to map each operation they take on a concrete object directly into a comparable operation on symbolic objects. Here, children are not changing their environments. Their environments provide stable structures, which can guide the development of their ideas.

Learning in Quadrant 2 (Symbiotic Tuning) is ideal for some abilities. Reading, for example, occurs in a highly stable environment of textual conventions — words are left-to-right, punctuation delimits sentences, and so forth. For reading, we want students to become highly tuned to a specific and recurrent environment. For abilities that need to transfer from one situation to another, symbiotic tuning is not ideal. In fact, one of the concerns of learning in Quadrant 1 (Induction) is that it can yield an unwanted symbiotic relation once student understanding stabilizes. For example, over time, children become more efficient at using base-10 blocks, and they become dependent on the blocks for that efficiency. Experimental studies indicate that learning with base-10 blocks does not transfer very well to symbolic tasks. The children's understanding is symbiotic on the base-10 blocks, and it does not survive when the blocks are removed.

In Quadrant 3 (Repurposing), learners' ideas are stable but their environments are adaptable. When children have stable ideas, they can adapt physical environments to support their problem solving. For example, 10-year-olds are familiar with multiplication, but they cannot always do complex problems in their heads. Nevertheless, the children have a sufficient grasp of multiplication that they can use just about any physical material to help solve multiplication problems they cannot solve mentally (Martin and Schwartz 2005). When people have strong ideas, they can repurpose many environments to support their continued learning.

4.1.1 *The case for considering mutual adaptations in physical environments*

In Quadrant 4 (Mutual Adaptation) children and their manipulatives both change form. A series of studies can help clarify what this means and why it is theoretically important (Martin and Schwartz 2005). In the first study, 10-year-olds tried to solve four problems of the form "indicate 1/4 of these 8 pieces". Each child tried two problems using plastic pieces spread across a table. They could move the pieces. Each child also tried two problems that showed pictures of pieces on a sheet of paper. The children had a pencil and paper to draw on the paper. The exact same children were three times more successful when they could manipulate the pieces (regardless of the order in which they completed the tasks). The children did not have a plan for how to solve the problems. Instead, they pushed the pieces rather haphazardly, sometimes one at a time and sometimes in groups. By adapting their environment, they began to reinterpret the meaning of the pieces; they realized that instead of counting each piece, they could count groups of pieces. For example, they might sort out four equal groups of pieces, and pick one group, solving the problem of what 1/4 means.

In the second study, we wanted to confirm that the children were learning by adapting their environment. An alternative might be that the children already had an algorithm to solve the problems, and they used the environment to help off-load the cognitive burden of executing the algorithm in their head. We gave children tile pieces in unstructured or pre-structured configurations. Figure 2 provides an instance. If the children already had an algorithm for solving these problems, then the pre-organized pieces should make the problem trivial. Children also received manipulative or picture versions of the materials. Again, when children could move the pieces, they were more accurate, even compared to when the pieces were pre-organized into the penultimate answer in a picture.

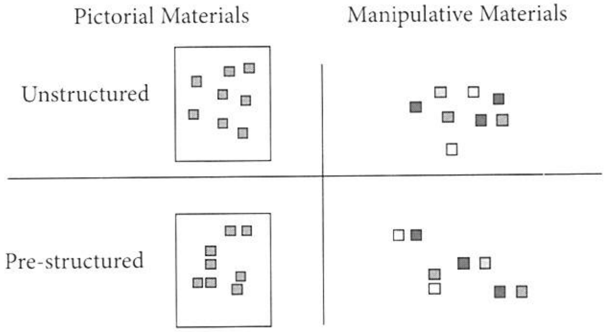


Figure 2. Unstructured and pre-structured materials.

The take home point of these examples is that adapting the environment was critical to children's own abilities to reinterpret the environment and solve these ratio problems. It seems like a theoretical mistake to treat the environment as fixed and only look at changes to the individual learners. The structure of the environment, from single units to groups, co-evolved with children's abilities to see groups in the environment.

4.1.2 *Predictions about the learning benefits of mutual adaptation in physical environments*

While the study of distributed learning requires an analysis of how individuals and their environments change with one another overtime, educators have a rightful concern for what individuals learn and can use once they leave a specific context. Thus, it becomes important to identify which quadrants of distributed learning are likely to yield individual learning that can transfer to new contexts. Our hypothesis is that mutual adaptation in Quadrant 4 is particularly beneficial for transfer. As people adapt their environments, they can learn which structures are adaptive and central to completing a given task. The hypothesized consequence is that they will be able to find or impose these structures in new environments that differ from the original context of learning.

Learning in Quadrant 4 may be contrasted with inductive learning in Quadrant 1, where the environment does not require any restructuring. To take the case of manipulatives in Quadrant 1, children induce how to rely on the given structure of the manipulative to solve problems. But, the children may never notice what structures in the environment made their manipulations effective. As an analogy, it might be something like using a calculator — one can learn what buttons to press, but this does not mean one will learn the structure of arithmetic that makes the button pressing effective. The consequence of this Quadrant 1 learning is that students may be less able to work in new environments that do not include all the structures built into their original environment.

To test our prediction that Quadrant 4 has special benefits for transfer compared to Quadrant 1, we conducted a teaching study with 9- and 10-year-olds. Over three days, children learned to solve problems like 1/4 + 1/2. Half

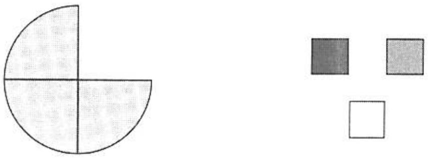


Figure 3. Pies indicate part–whole structure, whereas tiles require adaptation.

of the children used tiles and half used pie pieces. We chose these two types of materials because pie pieces have a part–whole structure built into their design. In Figure 3, one can readily see that the three pie-wedges do not make a whole. Tile materials do not manifest this part–whole structure, and it is much harder to see that the 3 tiles in the figure are parts of a whole. Thus, the pie pieces cause the learning to fall in Quadrant 1, where the environment has a stable structure, and students are expected to induce and use this structure. In contrast, the tile pieces require Quadrant 4 learning, because they do not manifest a part–whole structure. Children have to learn to adapt them into a part–whole structure to solve the problems. We thought this mutual adaptation would have benefits when the children had to transfer their learning to new environments. They would have a more complete interpretation of the structure of physical fractions, and they would be able to repurpose new environments into a useful structure.

The results indicated that both groups learned to do the fraction problems with their original learning material equally well. The difference between the two groups of children appeared when they had to transfer their embryonic understanding of fractions to new materials. At the end of each of the three days, we asked the children to solve similar problems to the ones they had solved earlier, except we gave them new materials (e.g., beans, fraction bars). The Tile students were more successful on the transfer problems than the Pie students were. They were able to figure out how to use the new materials to help them solve the problems.

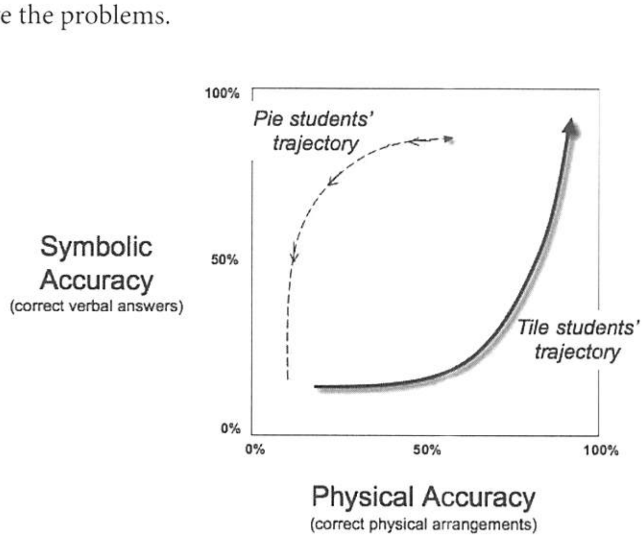


Figure 4. Learning trajectories as a function of physical and symbolic adaptations (Adapted from Martin and Schwartz 2005).

To describe this transfer effect, we found it valuable to consider adaptations to the environment and the children's symbolic responses. Figure 4 shows a schematic of the results. The vertical dimension indicates student accuracy at giving the correct symbolic answer to a problem. The horizontal dimension indicates whether the materials were accurately grouped for the problem. The arrows indicate the average trajectory of learning over the three days. Analyzing how the environment and the children's symbolic responses co-evolved reveals some interesting patterns. First, the Tile students were more likely than the Pie students to reach the top right corner of perfect performance. Second, the Tile students had a much more stable trajectory, whereas the Pie students often regressed from one day to the next. Third, the Tile children correctly adapted the physical materials slightly ahead of their abilities to give correct symbolic answers. In contrast, the Pie children tended to give correct symbolic answers before they could make correct physical adaptations. As the Tile students were more successful than the Pie students by the end, this result suggests that adaptations to the environment can lead to symbolic learning that has positive benefits for student understanding.

In summary, children learned to solve problems using their physical environment that they could not do in their heads (or with paper and pencil). Children depended on making structural changes to their environment to support learning. It was the opportunity to adapt their environment in the Tile condition that led to their subsequent abilities to adapt new environments. Had we ignored how the children distributed their learning with the environment, we would have no evidence as to why Pies were less useful than Tiles. However, by considering how children distributed their learning to the materials, it appears that the opportunity and ability to adapt the environment led to superior student learning that followed a more stable trajectory and could move from environment to environment.

4.1.3 *Technological application of mutual adaptation in a physical domain*

Technology can enhance the benefits of mutual adaptation. We created virtual manipulatives that children move like real manipulatives, for example, one-by-one or in groups. There are many virtual manipulatives available on-line, and they can provide many of the same benefits as real manipulatives do (Martin and Lukong, 2005). Our improvement has been to include a history feature. Children's behaviors are maintained in a database, and the children (and teacher) can replay their actions and see their answers. The value of this replay is that children can reflect on their adaptations to the environment.

The addition of a replay feature makes it easier for children to notice how the environment has been changing, and by hypothesis, they will learn more from their adaptations. For example, a first grader used our virtual manipulatives to solve the problem, "Sally has 5 apples. She eats some of them. Then she has 2 left. How many apples did she eat?". This problem uses an unknown intermediate quantity, which is harder for children than problems that use a final quantity as the unknown. The student made a collection of 5 pieces on the computer screen, then she removed pieces until there were two left. Next, she replayed her actions and counted the pieces that disappeared to get the answer "3". Being able to see her adaptations helped her decompose the task. She distributed her memory to the tool, so she could focus on getting to the target of two apples, and afterwards, she counted the number of apples she had removed. For students, the opportunity to observe their adaptations in process (and not just as finished products) can help them develop novel and appropriate interpretations of their actions sooner.

4.2 Socially distributed learning

We can extend the four quadrants of distributed learning to social interaction. For expository purposes, we will restrict our analysis to face-to-face interaction. The environment in this case is another person. Examples of Inductive learning might be an infant who realizes that his mother appears each time he cries in pain, or a student who learns that he is about to be scolded when his teacher begins with "Mr.". An example of Quadrant 2 (Repurposing) might be a mother who uses the infant's crying to tire him out, or a teacher who trains her students to be attentive when she says "Mr.".

The Symbiotic quadrant refers to situations where individuals refine stable patterns of interaction with one another. For example, partners may develop a particularly shorthand method of reference or emotional support, and they may learn to apportion chores in ways that are efficient given the peculiar structure of their lives. Symbiotic learning like this does not travel very well to new situations involving new people, though people do often try to impose a set of stable behaviors into new situations with varying degrees of success.

4.2.1 *The case for considering mutual adaptation in social environments*

In Quadrant 4, individuals adapt with one another and what emerges can be different than what either individual would learn without mutual adaptations. In these situations, it is difficult to describe learning by focusing solely on individual cognition. For example, we asked 15-year-olds to solve simple problems about gear movement such as, "Imagine there are four gears in a row that are touching like a row of quarters. If you try to turn the gear on the left end clockwise, what will happen to the gear on the right end?" We investigated whether the children spontaneously learned a parity rule — the first and last gears turn the same direction for an odd number of gears (Schwartz 1995). The children either worked alone or in pairs. The individuals learned the rule 14% of the time. Given this base rate, we can estimate how well the dyads should do if they communicate what they know perfectly and without "process loss". We do this by estimating the percentage of pairs that should include an individual who would have induced the rule working alone, and we assume this individual can convince his or her partner to accept the rule. This estimate is 26.5%. In fact, the pairs learned the parity rule 58% of the time. It is difficult to explain this level of learning in terms of individual cognition given that an analysis of individuals only predicts, at best, that 26.5% of the groups would learn the parity rule.

The members of the pairs somehow changed one another so they became more than the sum of two individual heads. This provides a parallel to the example of the young children who could only learn the meaning of "1/4 of 8" by pushing the pieces around. In both cases, the children's learning depended on an interaction with an environment that they were simultaneously modifying.

4.2.2 *Evidence on the learning benefits of mutual social adaptation*

We think many cognitive analyses of learning through social interaction have neglected the significance of mutual adaptation. A case in point involves the research on learning-by-teaching, where one student teaches another student and learns in the process of teaching. Learning-by-teaching should be an ideal instance of mutual adaptation, because the teacher learns by teaching, and the pupil learns by being taught. Both are adapting.

Even so, the cognitive explanations for the benefits of teaching have typically been reduced to individual cognitive processes. Chi et al. (1994: 441), for example, summarize the literature on learning-by-teaching:

> Citing 19 published studies... Webb found that giving elaborate explanations was positively related to individual achievement, whereas receiving elaborate explanations had few significant positive relationships with achievement... Such findings are consistent with the self-explanation effect because the advantages gained by explaining to others and to oneself are comparable.

While these findings are true as described, the authors have separated the learning of the teacher and the learning of the student into two separate, individual cognitive phenomena, giving and receiving. They do not treat teaching as a case of mutual adaptation, and we believe this oversight misses important causal elements in learning. For example, imagine a situation where teachers can teach and elaborate their ideas, but they never get to see how their students adapt in response. Would they learn so well?

Okita and Schwartz (2006) conducted a study to demonstrate that mutual adaptation is an important ingredient of learning-by-teaching. They showed that knocking out the opportunity for teachers to see how their student adapts reduces learning, even compared to teacher self-study. In the first phase of the experiment, college students read a passage on the mechanisms that cause people to heat up when they have a fever (e.g., a reduction in sweat production). They heard that they would be tutoring another student about the passage and should read the passage accordingly. The participants then taught Student X, a confederate, in a one-on-one session. Student X asked 5 questions, and the participants helped her with those questions. Student X followed a script to ensure she did not introduce any new information that would favor some participants over others.

In the next phase of the experiment, half of these participants received a new set of 5 questions. They were told that they would shortly be tutoring on the five questions and to consult the passage if they wanted. The other half of the participants saw Student X answer the same set of 5 new questions in an interview setting. The interviewer was another confederate who simply asked the questions without providing feedback. These participants saw a videotape of Student X answering the questions, and they were told that the video was a live feed. This way, every participant who observed Student X saw the same videotape, though they thought Student X's responses were her own adaptations of what they had taught her. Finally, all the participants took a surprise test of 5 new (and harder inference) questions.

Figure 5 shows how much the participants learned as measured by the final inference test. The participants who taught and observed Student X learned the most. Participants who taught Student X, but then did self-study on the second set of questions, were 27% less accurate. Evidently, the opportunity to passively observe how one's student adapts to what she has been taught is quite valuable for adapting one's own understanding — even more valuable than teaching and then doing it oneself. Moreover, we know that Student X did not introduce any special information when she answered questions in the faux-interview. A third group of students studied the first five questions in preparation for tutoring, and then they observed Student X answer the second set of five questions (as a "live feed" of a student answering questions). The figure shows these participants also did relatively poorly.

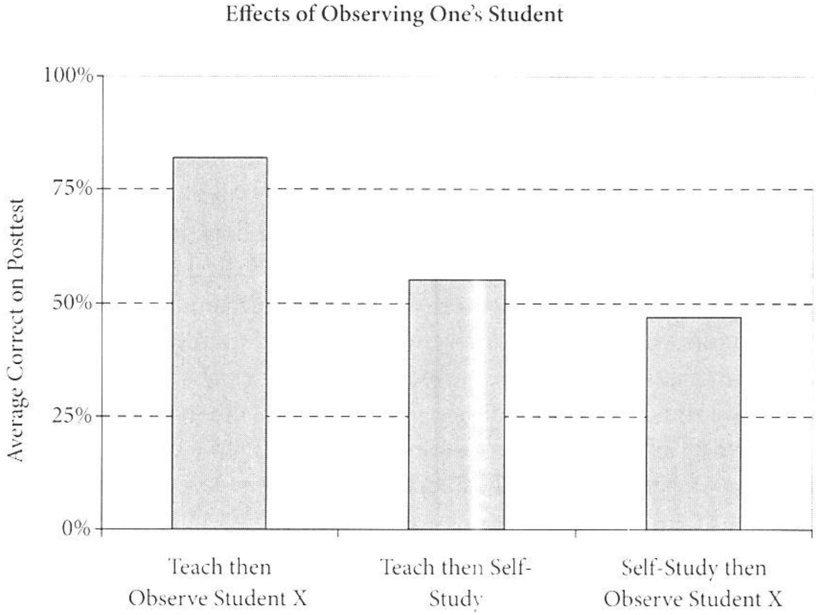


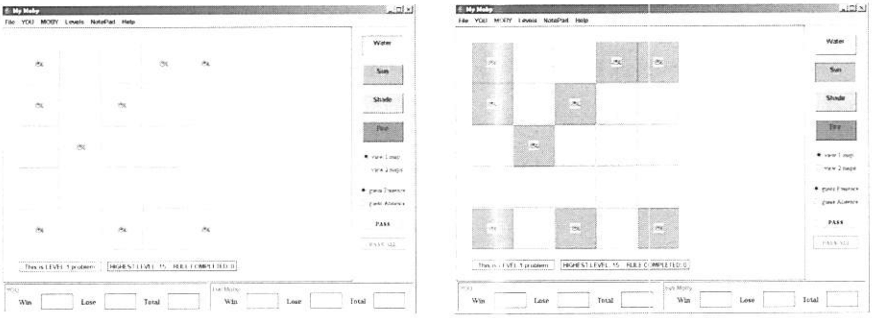
Figure 5. Participants learn more when they see their student answer questions (in uninformative ways) than when they answer the same questions themselves.

There are many useful results in this study. For example, the study provides a counter-example to the common wisdom that doing something oneself is better than observing. Observing can be extremely valuable in a context of mutual adaptation.

For our current purposes, however, the most important finding is that the participants who taught and observed their student learned the most. They learned more than the participants who taught but never observed how their student adapted their instruction. These latter participants also explained and elaborated their ideas when they taught. Therefore, the difference between the conditions cannot be reduced to the individual process of explaining or elaborating ones ideas. As fits our overall story, to make sense of a situation of mutual adaptation, like learning-by-teaching, it is useful to consider the overall system of distributed learning and not just individual processes.

4.2.3 *Implications of mutual social adaptation for technology*

We have transported the idea of mutual social adaptation into an educational technology called Teachable Agents (Schwartz et al., in press). Teachable Agents are computerized agents that students teach. Once they have been taught, the Teachable Agent can answer questions and solve problems using simple


My Moby
File YOU MOBY Levels NotePad Help
Water
Sun
Shade
Fire
PASS
This is LEVEL 1 problem
HIGHEST LEVEL 15 RULE COMPLETED 0
Win
Lose
Total

Overlay Water

Overlay Sun

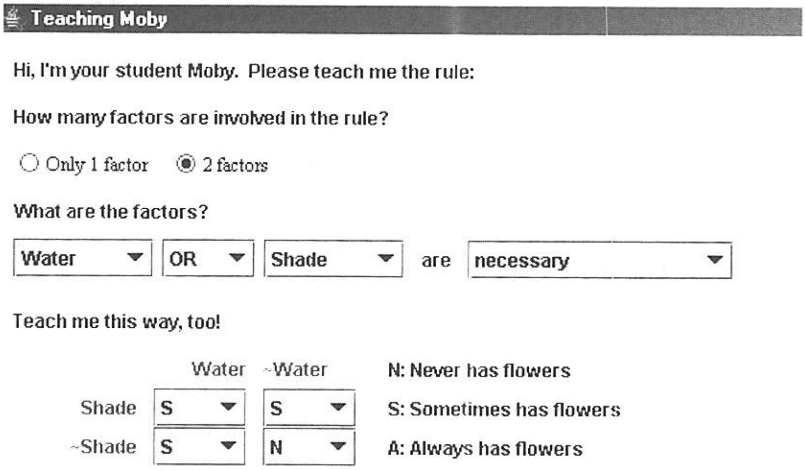
Figure 6. An interface for inducing how various factors influence a given outcome. Students can overlay factors (shown by colored cells) on a given outcome (shown by little flowers in the cells) to help induce the rule that regulates where the flowers will appear. Once they induce the rule, they teach a computer agent who then uses the rule to solve new problems.

artificial intelligence techniques. Students observe their agent's answers, and, depending on the quality of the answer, students can revise the agent's knowledge (and their own). Thus, Teachable Agents begin from the assumption that by adapting another social agent, students can adapt their own understanding. We have created several different classes of Teachable Agent to help students learn how to reason in different content areas (see <AAALab.Stanford.Edu>).

We conducted a study with a Teachable Agent, named Moby, to directly compare the inductive learning of Quadrant 1 with the mutual adaptation of Quadrant 4 (Schwartz et al., in press). Moby is a game context where students learn hypothetico-deductive reasoning. Students play a game that shows a given outcome (e.g., flowers growing). They have to induce the logical combination of factors that regulate the appearance of the outcome (e.g., "sun" is necessary and sufficient for flowers to appear).

Figure 6 provides an example of the interface for inducing a rule. Students overlay different factors on the grid by clicking on buttons that represent the factor. The right panel of Figure 6 shows what the interface looks like if students pick the correct factor for this particular problem. The rules can get quite complex; for example, "Sun and Not Water are necessary for the absence of the Flower".

Once players have induced a rule, they teach Moby by filling in the representations shown in Figure 7. Moby then plays a game based on the rule it has been taught; it tries to make predictions about where the flower will appear in a new grid (that follows the same underlying rule). Moby competes against "an



Hi, I'm your student Moby. Please teach me the rule:

How many factors are involved in the rule?

○ Only 1 factor ● 2 factors

What are the factors?

Water OR Shade are necessary

Teach me this way, too!

	Water	~Water
Shade	S	S
~Shade	S	N

N: Never has flowers
S: Sometimes has flowers
A: Always has flowers

Figure 7. Students teach their agent by using propositional and matrix formats.

evil agent" who also plays the prediction game. If Moby wins, students progress to the next level and a more difficult logical rule.

This particular Teachable Agent permits a nicely matched evaluation of induction (Quadrant 1) versus mutual adaptation (Quadrant 4). The study involved over a hundred high school seniors. Two conditions are of special relevance. In the Inductive condition, the students played the game themselves. In the Teach condition, the students saw their agent perform based on how it had been taught. Specifically, the Induction students induced rules, played against the evil agent to progress to the next level, and created a knowledge representation of each rule (per Figure 7). The Teach students also induced rules, created the same knowledge representation of the rule (to teach Moby), and then watched Moby play against the evil agent based on the rule. Thus, students in both conditions completed the same activities with the exception that the students in the Teach condition adapted an agent and watched its independent behavior, whereas the students in the Inductive condition used the rules themselves so there was no mutual adaptation.

Over the course of the study, there were no apparent differences between the conditions; both groups progressed through increasingly difficult problems at an equal rate of success. Students in both conditions adapted to the demands of the specific environment. However, as we have argued, the effects of mutual adaptation often show up in transfer situations when participants need to operate in a new environment. To test this claim, we gave the students a test of hypothetico-deductive reasoning a week later. The paper-and-pencil questions took several new forms that did not look like the original computer environment. Students in the Teach condition significantly outperformed the students in the Induction condition on every type of question. Thus, like the study with Student X, distributed learning can be more effective in helping students learn to handle new situations than simply doing things oneself.

5. Conclusion

In *The Lives of a Cell* (1978), Lewis Thomas points out the difficulty of determining the boundary of a living organism. Cells, for example, are often considered the fundamental unit of life. But, within every cell are mitochondria that produce the cell's usable energy. Mitochondria have a different genetic make-up from the cell, and they appear to be "separate" living entities with their own reproductive mechanisms. This is an excellent example of the inextricable interdependence of living organisms, and the difficulty of identifying any single "unit" of life. Thomas ultimately concludes that the earth itself should be considered a single organism.

Distributed cognition brings up similar issues, because it raises the question of how we should locate the "unit" of cognition. Harnad (2005), for example, rejects an information processing definition of cognition. Instead, he proposes that the unit of cognition should be identified phenomenologically by the introspective feeling of thought. In this solution, distributed cognition does not exist, because cognition can only be felt by individuals. He proposes "cooperative cognition" as a better way to think about dependencies between individuals and an intelligently designed environment.

In our own work, we have tried another approach to the boundary problem. We begin with the assumption of interdependence instead of singular cognition. Our question is how people and their environments adapt such that they become more or less interdependent over time. Our finding has been that people and their environments can be highly interdependent during initial learning. At the same time, certain types of interdependence in early learning, most notably mutual adaptation, can help prepare people to be *less* dependent on their immediate environment and more adaptive when they confront new environments.

Although we have made some headway in our analysis of distributed learning, there are some notable omissions. Perhaps the most glaring omission is that we have not offered a solid metric for evaluating the degree to which the elements of a distributed learning system are stable or amenable to adaptation. In our examples, we have relied on intuitively different situations, but this can only take us so far. At some point, it will be necessary to define which changes constitute an adaptation and which changes simply reflect the usual processes of a stable environment. This applies to both individuals and environments.

Cognitive research has made some headway in distinguishing incremental changes from larger adaptations in individual cognition (e.g., automaticity versus conceptual change; for a review, see Schwartz, Bransford, and Sears 2005). The work of Hatano and Inagaki (1986), for example, distinguishes routine expertise from adaptive expertise. Routine experts are highly efficient in their usual tasks, whereas adaptive experts seek out new challenges and sources of learning. We hope that cognitive science begins to see a similar value to making headway in distinguishing "stable" and "adaptive" environments. The appearance of adaptive expertise, for example, is surely a function of whether an environment allows for individual adaptation.

Note

* This material is based upon work supported by the National Science Foundation under Grants No. BCS-0214548, BCS-0214549, REC-0231946, and SLC-0354453. The work on virtual manipulatives was supported by the Center for Instructional Technology and the Office of the Vice President at the University of Texas, Austin. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the University of Texas.

References

Carpenter, T.P., Corbitt, M.K., Kepner, H.S., Jr., Lindquist, M.M., and Reys, R.E. 1980. "National assessment: A perspective of students' mastery of basic mathematics skills". In M.M. Lindquist (ed), *Selected Issues in Mathematics Education.* Reston, VA: National Council of Teachers of Mathematics, 215–227.

Chambers, D., and Reisberg, D. 1985. "Can mental images be ambiguous?". *Journal of Experimental Pychology: Human Perception and Performance* 11: 317–328.

Chi, MT.H., de Leeuw, N., Chiu, M-H., and LaVancher, C. 1994. "Eliciting self-explanations improves understanding". *Cognitive Science* 18: 439–477.

Dienes, Z.P. 1973. *The Six Stages in the Process of Learning Mathematics.* Slough: National Foundation for Education Research/Nelson.

Fuson, K.C. and Briars, D.J. 1990. "Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction". *Journal for Research in Mathematics Education* 21(3): 180–206.

Glenberg, A.M., Gutierrez, T., Levin, J.R., Japuntich, S., and Kaschak, M. 2004. "Activity and imagined activity can enhance young children's reading comprehension". *Journal of Educational Psychology* 96(3): 424–436.

Hall, N. 1998. "Concrete representations and the procedural analogy theory". *Journal of Mathematical Behavior* 17(1): 33–51.

Harnad, S. (2005) Distributed Processes, distributed cognizers and collaborative cognition. *Pragmatics & Cognition* 13(3): 501–514.

Hatano, G. and Inagaki, K. 1986. "Two courses of expertise". In H. Stevenson, H. Azuma, and K. Hakuta (eds), *Child Development and Education in Japan*. New York: W.H. Freeman, 262–272.

Hatano, G., and Osawa, K. 1983. "Digit memory of grand experts in abacus-derived mental calculation". *Cognition* 15: 95–110.

Hutchins, E. 1995. "How a cockpit remembers its speed". *Cognitive Science* 19: 265–288.

Kirsh, D. 1996. "Adapting the environment instead of oneself". *Adaptive Behavior* 4(3/4): 415–452.

Kerslake, D. 1986. *Fractions: Children's Strategies and Errors*. Berkshire: NFER/Nelson.

Martin, T. 2004. "Coevolution of symbolic and concrete dimensions of understanding". In *Proceedings of the 6th International Conference of the Learning Sciences*, Santa Monica, CA, 622.

Martin, T. and Lukong, A. 2005. "Virtual manipulatives: How effective are they and why?". Paper presented at the *American Educational Research Association Annual Conference*, Montreal.

Martin, T., and Schwartz, D. L. 2005. "Physically distributed learning: Adapting and reinterpreting physical environments in the development of the fraction concept". *Cognitive Science* 29: 587–625.

Norman, D.A. 1988. *The Psychology of EverydayThings*. New York: Basic Books.

Okita, S.Y. and Schwartz, D.L. 2006. "When observation beats doing: Learning by teaching". In S. Bards, K. Hay, D. Hickey (eds), *7th International Conference of the Learning Sciences* (ICLS), Vol. 1. Mahwah, NJ: Laurence Erlbaum, 509–516.

Pea, R.D. 1993. "Practices of distributed intelligence and designs for education". In G. Salomon (ed), *Distributed Cognitions*. New York: Cambridge University Press, 47–87.

Piaget, J. 1953. "How children form mathematical concepts". *Scientific American* 189(5): 74–79.

Schwartz, D. L. 1995. "The emergence of abstract representations in dyad problem solving". *Journal of the Learning Sciences* 4: 321–354.

Schwartz, D.L., Blair, K.P., Biswas, G., Leelawong, K., and Davis, J. In press. "Animations of thought: Interactivity in the teachable agents paradigm". In R. Lowe and W. Schnotz (eds), *Learning with Animation: Research and Implications for Design*. Cambridge: Cambridge University Press.

Schwartz, D.L., Bransford, J.D., Sears, D.L. 2005. "Efficiency and innovation in transfer". In J. Mestre (ed), *Transfer of Learning from a Modern Multidisciplinary Perspective*. Norwich, CT: Information Age Publishing, 1–54.

Thomas, L. 1978. *The Lives of a Cell*. Middlesex, UK: Penguin Books.

Zhang, J. and Norman, D.A. 1994. "Representation in distributed cognitive tasks". *Cognitive Science* 18: 87–122.

Authors' addresses

Daniel L. Schwartz
485 Lasuen Mall
School of Education
Stanford University
Stanford, CA 94305
U.S.A.

Daniel.schwartz@stanford.edu

Taylor Martin
Science and Math Education Program
1 University Station D5705
University of Texas, Austin
Austin, TX 78712-0382
U.S.A.

taylormartin@mail.utexas.edu

About the authors

Daniel L. Schwartz is Associate Professor of the Learning Sciences in the School of Education at Stanford University. He received his Ph.D. in Human Cognition from Columbia University. His work combines cognitive science and classroom research to understand processes of learning and problem solving and to help design and evaluate novel forms of instruction using advanced technologies.

Taylor Martin is Assistant Professor of Science and Mathematics Education at the University of Texas at Austin. She received a B. A. in Linguistics from Dartmouth College, an M.S. in Psychology from Vanderbilt University, and a Ph.D. in Education from Stanford University. Her primary research interest is how people learn content in complex domains from active participation. She examines how hands-on activities impact mathematics learning and investigates the development of adaptive expertise.

Distributed cognition, representation, and affordance

Jiajie Zhang and Vimla L. Patel
University of Texas at Houston / Columbia University

This article describes a representation-based framework of distributed cognition. This framework considers distributed cognition as a cognitive system whose structures and processes are distributed between internal and external representations, across a group of individuals, and across space and time. The major issue for distributed research, under this framework, are the distribution, transformation, and propagation of information across the components of the distributed cognitive system and how they affect the performance of the system as a whole. To demonstrate the value of this representation-based approach, the framework was used to describe and explain an important, challenging, and controversial issue — the concept of affordance.

Keywords: distributed cognition, internal and external representations, affordance

1. Introduction

Distributed cognition is a scientific discipline that is concerned with how cognitive activity is distributed across internal human minds, external cognitive artifacts, and groups of people, and how it is distributed across space and time (Hutchins 1995a, 1995b; Norman 1991; Zhang 1997a, 1997b, 1998; Zhang and Norman 1994). In this view, people's intelligent behavior results from interactions with external cognitive artifacts and with other people, and people's activities in concrete situations are guided, constrained, and to some extent, determined by the physical, cultural, and social contexts in which they are situated (Clancey1997; Suchman 1987). The unit of analysis for distributed cognition is a distributed cognitive system composed of a group of people interacting with external cognitive artifacts. Such a distributed system (e.g., the cockpit of a commercial airplane) can have cognitive properties that differ radically

Pragmatics & Cognition 14:2 (2006), 333–341.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

from the cognitive properties of the components, and these properties cannot be inferred from the properties of the components alone, no matter how much we know about the details of the properties of those components (Hutchins, 1995b). In general terms, we describe the components of a distributed cognitive system as internal and external representations. Internal representations are the knowledge and structure in individuals' minds; and external representations are the knowledge and structure in the external environment (Zhang 1997b; Zhang and Norman 1994).

The term "distributed cognition" has been used in different ways. In our own research and in this paper, we use it to refer to the scientific discipline that is concerned with the distribution of information and knowledge between and across internal and external representations. It does not necessarily mean "distributed knowing or distributed thinking" of a distributed system. We do not argue for or against the assertion that a distributed system has consciousness or awareness or the assertion that a distributed system can think or reason in the same way that an individual mind does. This is a philosophical question, one insightful and deep analysis of which was offered by Harnad (2005). For the purpose of studying the behavior of a distributed cognitive system it is usually sufficient to understand how information and knowledge are distributed and propagated across the various components of the distributed system.

In this paper, we first describe two types of distributed cognition: between an individual mind and an external artifact and between individual minds. Then we will show how distributed cognition can be used to offer a novel solution to a difficult and challenging issue: affordance.

2. Distributed cognition between individuals and artifacts

A wide variety of complex information processing tasks require the processing of information distributed across internal minds and external artifacts. It is the interwoven processing of internal and external information that generates much of a person's intelligent behavior. Let us consider multiplying 965 by 273 using paper and pencil. The internal representations are the meanings of individual symbols (e.g., the numerical value of the arbitrary symbol "5" is five), the addition and multiplication tables, arithmetic procedures, etc., which have to be retrieved from memory. The external representations are the shapes and positions of the symbols, the spatial relations of partial products, etc., which can be perceptually inspected from the environment. To perform this task, people need to process the information perceived from external representations and the information retrieved from internal representations in an interwoven, integrative, and dynamic manner. Zhang and Norman (1994) developed a framework of distributed representations to account for the behavior in these types of distributed cognitive tasks.

One important aspect emphasized by distributed cognition research is that external representations are more than inputs and stimuli to the internal mind. External representations have many non-trivial properties. For many tasks, external representations are intrinsic components, without which the tasks either cease to exist or completely change in nature. Zhang (1997b) reviewed and summarized the following properties of external representations:

- Provide short-term or long-term memory aids so that memory load can be reduced.
- Provide information that can be directly perceived and used such that little effortful processing is needed to interpret and formulate the information explicitly.
- Provide knowledge and skills that are unavailable from internal representations.
- Support perceptual operators that can recognize features easily and make inferences directly.
- Anchor and structure cognitive behavior without conscious awareness.
- Change the nature of a task by generating more efficient action sequences.
- Stop time and support perceptual rehearsal to make invisible and transient information visible and sustainable.
- Aid processibility by limiting abstraction.
- Determine decision making strategies through accuracy maximization and effort minimization.

3. Distributed cognition across individuals

Cognition can also be distributed across a group of individuals. For this type of distributed cognition, there are two different views. The reductionist view considers that the cognitive properties of a group can be entirely determined by the properties of individuals. In this view, to understand group behavior, all we need is to understand the properties of individuals. In contrast, the interactionist view considers that the interactions among the individuals can produce emergent group properties that cannot be reduced to the properties of the individuals. In this view, to study group behavior, we need to examine not only the properties of individuals but also the interactions among the individuals. Examples of emergent group properties include group affect (George 1990), collective efficacy (Bandura 1986), and transactive memory systems (Wegner 1987).

One important issue in distributed cognition across a group of individuals is the group effectiveness problem (Foushee and Helmreich 1988). A group of minds can be better than one (process gain) because in a group there are much more resources, task load and memory load are shared and distributed, errors are cross-checked, and so on. The performance of a group can also be worse than that of an individual (process loss) because in a group communication takes time, knowledge may not be shared and different strategies may be used by different individuals. This phenomenon has been shown in a clinical environment where people work face to face, sharing Tacit knowledge (Patel *et al.* 2000) and at a distance in an internet based co-laboratory (Patel et al. 1999). Using the framework of distributed representations developed by Zhang and Norman (1994), Zhang (1998) demonstrated empirically that whether two minds were better or worse than one mind depended on how the knowledge was distributed across the two minds. The issue of group effectiveness is especially important in healthcare and has received some attention recently (Patel et al. 1996).

4. Affordance as distributed dognition

Gibson's (1977, 1979) "affordance" is an intriguing, useful, but controversial concept. According to Gibson, the environment not only serves as the surfaces that separate substances from the medium in which the animals live, but also affords animals in terms of terrain, shelters, water, fire, objects, tools, animals, human displays, etc.; and there is not only information for the perception of the environment, but also information for the perception of what the environment affords. Gibson's affordance has the following properties:

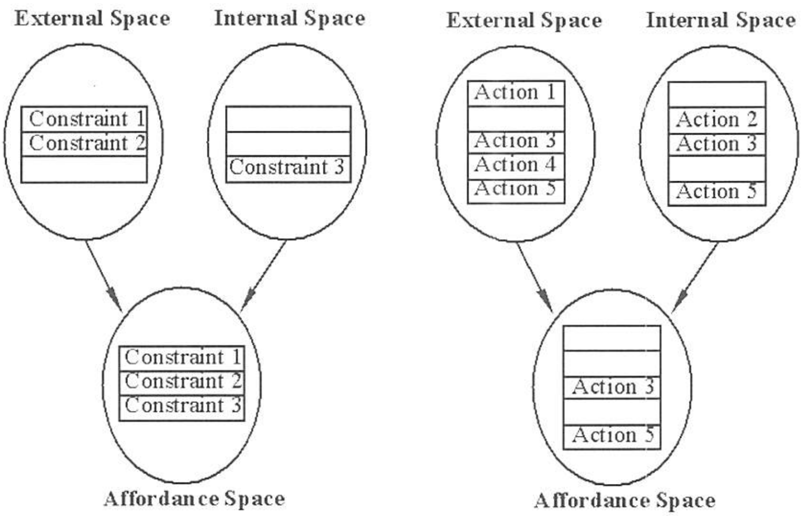
- Affordances provided by the environment are what it offers, what it provides, what it furnishes, and what it invites.
- The "values" and "meanings" of things in the environment can be directly perceived. The "values" and "meanings" are external to the perceiver.
- Affordances are relative to animals. They can only be measured in ecology, but not in physics.
- An affordance is an invariant.

- Affordances are holistic. What we perceive when we look at objects are their affordances, not their dimensions and properties.
- An affordance implies complementarity of the perceiver and the environment. It is neither an objective property nor a subjective property, and at the same time it is both. It cuts across the dichotomy of subjective-objective. Affordances only make sense from a system point of view.

Gibson's affordances capture a fundamental aspect of human perception and cognition, that is, the fact that much information needed for perception and action is in the environment as invariants which can be picked up directly. Gibson's notion of affordances was developed primarily in the study of visual perception, and his affordances are basically for environmental things such as substances, media, layouts, events, etc. Gibson hinted that affordances are not only for visual perception, but also could be for biological (e.g., toxin and nutrition) and cultural (e.g., mail box) processes. However, he did not fully explore these implications. In this section, we use the concept of distributed cognition to develop a theoretical framework of affordances.

Affordances are the allowable actions specified by the environment coupled with the properties of the organism. In distributed cognition, affordances can be considered as distributed representations extended across the environment and the organism. The structures and information in the environment specify the external representation space. The physical structures of the organism and the structures and mechanisms of internal biological, perceptual, and cognitive faculties specify the internal representation space. The external and internal representations together specify the distributed representation space, which is the affordance space (Figure 1). The external and internal representation spaces can be described by either constraints or allowable actions. Constraints are the negations of allowable actions. That is, the allowable actions are those satisfying the constraints, and the constraints set the range of the allowable actions. If the external and internal representation spaces are described by constraints, then the affordances are the disjunction of the constraints of the two spaces. If the external and internal representation spaces are described by allowable actions, then the affordances are the conjunction of the allowable actions of the two spaces.

Gibson's original affordances are basically those that are specified by the relations between the physical structures of the environment and the physique of the organism (e.g., chairs afford sitting for people). Two of the fundamental properties of affordances are the complementarity of the environment and the organism and the direct, effortless pickup of affordances. These two properties


External Space
Internal Space
Constraint 1
Constraint 2
Constraint 3
Constraint 1
Constraint 2
Constraint 3
Affordance Space
External Space
Internal Space
Action 1
Action 3
Action 4
Action 5
Action 2
Action 3
Action 5
Action 3
Action 5
Affordance Space

(A) Constraints
Affordance = Internal $\cup$ External

(B) Allowable Actions
Affordance = Internal $\cap$ External

Figure 1. The distributed cognition framework of affordances. (A) The representation spaces are described by constraints. The affordance space is the *disjunction* of the external and internal representation spaces. (B) The representation spaces are described by possible actions. The affordance space is the *conjunction* of the external and internal representation spaces.

exist not only at the level of organism's body structures, but also at many other levels. The distributed cognition framework of affordance can extend Gibson's original affordances to phenomena at other levels.

Under the distributed cognition framework, affordances are distributed representations extended across external (the environment) and internal (the organism) representations. External representations belong to the environment; and internal representations belong to the organism. External representations can be at the levels of chemical processes, physical configurations, spatio-temporal layouts, and symbolic structures, which correspond to the levels for internal representations: biological mechanisms inside the body, the physique of the organism, perceptual systems, and cognitive structures and processes.

This framework can generate a categorization of affordances:

- *Biological Affordance.* Biological affordance is based on biological processes. For example, a healthy mushroom affords nutrition, while a toxic mushroom affords dying. This is at the level of biology.

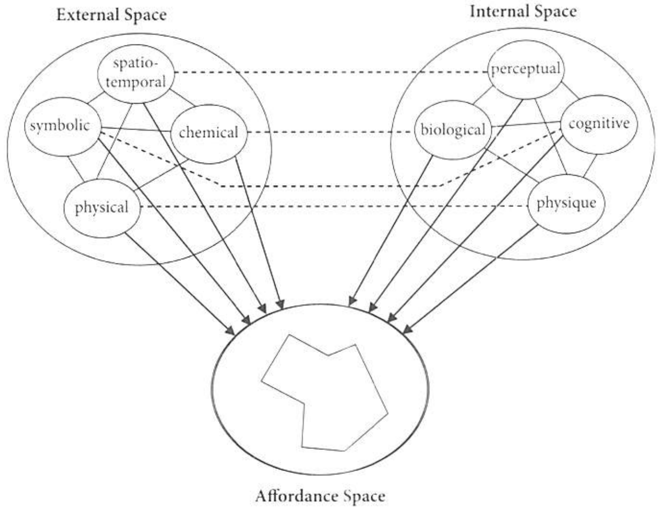


Figure 2. A categorization of affordances from the perspective of distributed cognition.

- *Physical Affordance.* Physical affordance is for tasks that are mainly constrained by physical structures. For example, the flat horizontal panel on a door can only be pushed. As another example, an open environment affords locomotion in any direction over the ground, whereas a cluttered environment affords locomotion only at opening.
- *Perceptual Affordance.* In this category, affordances are mainly provided by spatial mappings. For example, if the switches of the stovetop burners have the same spatial layout as the burners themselves, the switches provide affordances for controlling the burners. Another example is the pictorial signs for ladies' and men's restrooms in airports.
- *Cognitive Affordance.* Affordances of this type are provided by cultural conventions. For example, for traffic lights, red means "stop", yellow means "prepare to stop", and green means "go".
- *Mixed Affordance.* Many affordances are provided by a combination of more than one module. For example, shoelace affords tying shoes. This affordance is a conjunction of physical affordance and cognitive affordance: the physical property of shoelace and the knowledge of how to make a tie. The "mailbox" example given by Gibson is also a mixed affordance. A mailbox does not provide the affordance of mailing letters for a person who has no knowledge about postal systems. In this case, knowledge (cognitive affordance) and structure of a mailbox (physical affordance) are both involved in constructing the affordance for mailing and receiving letters.

5. Conclusion

Distributed cognition, in our view, is a term for a branch of cognitive science that is concerned with a special type of cognitive systems whose structures and processes are distributed between internal minds and external environment, across a group of individual minds, and across space and time. From the distribute cognition perspective, the unit of analysis is the interaction between the components of the system, not the components themselves. We applied the distributed cognition framework to address an important, challenging, and controversial issue — the concept of affordance originally proposed by Gibson.

References

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Clancey, W.J. 1997. *Situated Cognition: On Human Knowledge and Computer Representations.* Hillsdale, NJ: Erlbaum.

George, J.M. 1990. "Personality, affect, and behavior in groups". *Journal of Applied Psychology* 75: 107–116.

Foushee, H.C. and Helmreich, R.L. 1988. "Group interaction and flight crew performance". In E.L. Wiener and D.C. Nagel (eds), *Human Factors in Aviation.* Orlando, FL: Academic Press, 189–227.

Gibson, J.J. 1977. "The theory of affordances". In R.E. Shaw and J. Bransford (eds), *Perceiving, Acting, and Knowing.* Hillsdale, NJ: Lawrence Erlbaum, 67–82.

Gibson, J.J. 1979. *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

Harnad, S. 2005. "Distributed processes, distributed cognizers, and collaborative cognition". *Pragmatics and Cognition* 13(3): 501–514.

Hutchins, E. 1995a. *Cognition in the Wild.* Cambridge, MA: The MIT Press.

Hutchins, E. 1995b. "How a cockpit remembers its speed". *Cognitive Science* 19: 265–288.

Norman, D.A. 1991. "Cognitive artifacts". In J.M. Carroll (ed), *Designing Interaction: Psychology at the Human-Computer Interface.* New York: Cambridge University Press, 17–38.

Patel, V.L., Cytryn, K.N., Shortliffe, E.H., and Safran, C. 2000. "The collaborative health care team: The role of individual and group expertise". *Teaching and Learning in Medicine* 12(3): 117–132.

Patel, V.L., Kaufman, D.R., Allen, V.G., Shortliffe, E.H., Cimino, J.J., and Greenes, R.A. 1999. "Toward a framework for computer-mediated collaborative design in medical informatics". *Methods of Information in Medicine* 38: 158–176.

Patel, V.L., Kaufman, D.R., and Magder, S. 1996. "The acquisition of medical expertise in complex dynamic environments". In A. Ericsson (ed), *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games.* Hillsdale, NJ: Lawrence Erlbaum, 127–165.

Suchman, L. 1987. *Plans and situated actions: The problem of human–machine communication*. New York: Cambridge University Press.

Wegner, D.M. 1987. "Transactive memory: A contemporary analysis of the group mind". In B. Mullen and G.R. Goethals (eds), *Theories of group behavior*. New York: Springer, 185–208.

Zhang, J. 1997a. "Distributed representation as a principle for the analysis of cockpit information displays". *International Journal of Aviation Psychology* 7: 105–121.

Zhang, J. 1997b. "The nature of external representations in problem solving". *Cognitive Science* 21: 179–217.

Zhang, J. 1998. "A distributed representation approach to group problem solving". *Journal of the American Society of Information Science* 49(9): 801–809.

Zhang, J. and Norman, D.A. 1994. "Representations in distributed cognitive tasks". *Cognitive Science* 18(1): 87–122.

Authors' addresses

Jiajie Zhang
School of Health Information Sciences
University of Texas, Houston
7000 Fannin, Ste. 600
Houston, TX 77030
U.S.A.

jiajie.zhang@uth.tmc.edu

Vimla L. Patel
Department of Biomedical Informatics
Columbia Presbyterian Medical Center
Vanderbilt Clinic 5
622 West 168th St.
NewYork, NY 10032-3720
U.S.A.

patel@dmi.columbia.edu

About the authors

Jiajie Zhang is a Professor and the Associate Dean for Research in the School of Health Information Sciences at the University of Texas at Houston. His research interests include human–computer interaction, cognitive science and engineering, distributed cognition, and health informatics.

Vimla L. Patel is a Professor of medical informatics and medical psychology at Columbia University. Her research interests include medical cognition, cognitive evaluation of human–computer interaction, collaborative healthcare teams, and naturalistic decision making in critical-care environments.

Categorization and technology innovation

Jeffrey M. Stibel
Web.com, Inc.

Theories on categorization have led to numerous technical innovations. Starting with artificial intelligence and neural models, scientists have leveraged psychological theories to drive forward innovative technology. More recently, software companies and Internet firms have implemented high technology software developed from cognitive theory. One class of systems rooted in the philosophical tradition stresses the importance of explanation and function. Another focuses on feature similarity and rule-based reasoning. Both approaches have had modest success and solve fundamental problems, but neither has achieved the higher-levels of categorization found in humans. The current paper critically analyzes the various theories of categorization and their collective impact on recent technology innovations.

Keywords: categorization, concepts, essentialism, technology, category modeling, innovation

1. Introduction

Since Socrates, philosophers have struggled with how things are categorized, organized and embedded in the world and stored in our minds. The ancient Greeks originally thought that categories were composed of a few primary elements (fire, earth, air, and water). When Plato put fourth his theory of forms, the modern day field of concepts and categories was born. It was not until Aristotle, however, that the metaphysics of philosophy and religion were removed and a more pragmatic science of categorization emerged. Philosophical theories of categorization have been extended to include notions of explanation, such as essentialism (Putnam 1988). Essentialism argues that category formation is driven by fundamental essences that bind all categories. For example, the ability to transport people by water is an essential characteristic of boats. This explanation adequately defines the category of boats. Essentialism does

Pragmatics & Cognition 14:2 (2006), 343–355.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

a good job accounting for man-made artifacts but cannot account for many living things or other natural kinds. A number of recent technologies have utilized essentialism to further innovation, such as Internet search engines, which extract an essence using user-generated keywords upon which the search engine bases its results.

Research on categorization has also led to theories based more on probability. Probabilistic accounts of categorization have led to a number of theories with rigid boundaries as well as less defined versions, such as family resemblance and prototype theory. Probabilistic theories of categorization have been used extensively in software development; most notably in creating object oriented programming and Internet recommendation engines.

Modern advances in technology have enabled scientists to model psychological states such as categorization. These models have advanced our understanding significantly. Through the use of such models, various psychological and philosophical theories have been tested and extended. One prescient finding is that models of the categorization process based on the above theories often can only account for a small set of categories, typically within one domain. This has led to hybrid approaches that try to incorporate both probability and explanation. Technology innovation, to no surprise, has followed this development.

The present paper reviews various theories of categorization based on essentialism and probability, while examining their impact on technology innovation. Using relevant current examples, we outline how these theories of categorization have played a role in developing technologies. We cite a number of recent examples of technology innovations that were driven in large part by theories of categorization.

2. Psychological models of categorization and the technologies they enable

One of the most enduring theories of categorization was derived from Aristotle. Aristotle believed that categorization is driven by certain necessary and sufficient conditions. He argued that every object has features that define rigid category boundaries. For example, to be considered a square, an object must have exactly four parallel sides and four right angles. In this way, all members of a category are equally good examples of that category.

Psychologists in the mid 20th century revived this "classical" approach and tested it experimentally. The classical view is similar to Aristotle's original approach in that certain necessary and sufficient features make up rigid category boundaries. These psychologists postulate a three-fold theory: that category descriptors contain "summary representations" abstracted from, and descriptive of, its contents; that these descriptors are "singly necessary…and jointly sufficient to define that concept"; and that category descriptors are hierarchical, wherein sub-categories must possess all of the descriptive features of their super-ordinates (Smith and Medin 1981).

Technological modeling demonstrated a number of weaknesses in the classical approach to categorization. Early artificial intelligence, for example, was based on the classical model, as was simple analog robotics. Over time, these models incorporated heuristics that led to more flexible rule-based approaches. These early models focused on building rule-based systems that bound categories by necessary and sufficient conditions. The early artificial intelligence attempts provided moderate success but were strictly limited by technological deficiencies (i.e., not enough memory, power, and speed). In addition, they demonstrated a number of weaknesses of the classical approach.

While the classical approach appeals to common sense, it fails to broadly account for categorization. Paramount among its limitations, the classical view fails to account for cases of exclusive disjunction in concept formation (Smith and Medin 1981). Because the classical approach insists on necessary and sufficient conditions of categorization, features of objects must always be present. Categories such as "automobile", where it is possible to be a member by being a car or a truck but not both, do not fit well into the classical approach. Because of this, classical artificial intelligence models cannot categorize objects whose features are more similar to those of one category yet are members of another (e.g., whales are more similar to fish, yet are considered mammals). Another reason why the classical approach fails is that it separates objects into well-defined categories that are linearly separable. An early neural network model, the Perceptron, used feature weights to create linearly defined categories (Anderson 1995). Of course, its fate and demise was similar to the classical approach (for a devastating attack, see Minsky and Papert 1988).

Nevertheless, the classical model has been used to advance technology. Early versions of problem solving, decoding and decryption software employed techniques derived from generating clearly defined boundaries. Given the model's limits, it should come as no surprise that these programs were not particularly successful at solving anything other than the most rudimentary problems.

Early categorization of content on the Internet also utilized classical models. Yellow pages providers, such as SuperPages, InfoSpace, and YellowBook, in an attempt to resolve user frustration with navigating through business listings online, successfully leveraged linear category hierarchies (Stibel 2005). Search engines, such as Yahoo!, Lycos, and AltaVista, built category trees or directories rooted in the classical model. The Open Directory, developed by Netscape and AOL, utilized a collaborative environment to classify and store Internet content in linear categories. While successful in certain circumstances, the limits of this approach gave way to more sophisticated implementations.

The failure of classical models to universally account for concept formation led to a series of probabilistic approaches. These theories hold that features are neither necessary nor sufficient for categorization. Rather, categories are made up of objects that share some number of similar features. In that way, an object is judged to be a category member if it shares some proportion of typical features with other category members. Depending on how the approach is implemented, probabilistic approaches can account for a number of problems with the classical view that were discovered experimentally: non-linearity (Minsky and Papert 1988), fuzzy boundaries (Rosch and Mervis 1975), and hierarchical organization (Collins and Loftus 1975). While a vast number of probabilistic approaches exist, they broadly fit into three types.

In family resemblance theory, categories are determined through a hierarchy of similar features. Features that a higher-level category possesses will be passed on, for the most part, to all of its sub-categories. According to Smith and Medin (1981: 45), "an item is a typical subset or member of a concept if it contains features shared by many other members of that same category".

This approach adequately accounts for non-linearities and fuzzy category boundaries, while retaining the notion of feature similarity among categories. Rosch and Mervis (1975) demonstrated experimentally that people tend to use the totality of feature similarity in their category judgements. They found, for example, that subjects base their categorization decisions of "furniture" on the totality of similar features (i.e., chair, sofa, cushion, rug, vase, telephone) — where some objects are more representative than others.

Family resemblance theory provided an analog for object oriented programming languages, such as Java and C++. Object oriented programming uses a probabilistic structure to map sub-programs to a master hierarchy of code that can be utilized and interpreted by multiple programmers. Because programming languages such as Java and C++ mimic storage in the mind, they enable sharing amongst programmers. This innovation, derived from a family-resemblance approach, redefined the theory of computing and enabled a revolution in software.

The family resemblance model was extended by a second probabilistic approach — prototype theory. Prototype theory holds that a single category member is chosen as the most representative of its category. Novel objects are then compared to this prototype and, depending on how many features the two have in common, a category decision is made. For example, people create a prototypical chair based on observations of a variety of different chairs. They then use this example to compare features of novel items to the prototype.

Good examples of prototype theory in action are those used with digital video recorders (DVRs), such as Tivo. Tivo has programs designed to generate viewing recommendations for users. Tivo employs a prototype model where users rate certain shows and an algorithm compares that prototype to other television shows. The closer the match, the more likely it is that the viewer will enjoy the show and Tivo will make a recommendation based on that similarity rating. Early Internet recommendation engines, such as mySimon and Amazon.com, utilized similar models. They analyzed prototypical buying behavior to predict future purchases with moderate success.

The problem with prototype theories is that they tend to eliminate a large amount of data not represented by the prototype. This has the benefit of being very easy to model, as information can be stored in low-memory systems. However, as early neural networks demonstrated, prototype models could not extrapolate to other categories because of incomplete information. Kruschke's ALCOVE model (1992), which was principally based on another probability model, demonstrated the failure of prototyping as a broad based application. ALCOVE showed that a single prototype was not sufficient to categorize less similar members of a category, such as whales (mammals) and penguins (birds). As a result, prototype theories cannot account for broad category variability (Fried and Holyoak 1984).

As a better account for category variability, Medin and Shaffer (1978) proposed a theory based on multiple category exemplars. Similar to prototype theory, exemplar models assume that novel objects activate learned exemplars of a particular category. However, unlike prototype theory, exemplar models allow for multiple prototypes in any given category. People use the most similar exemplar(s) to a particular novel object in order to determine category membership. For example, a "chow" may elicit the exemplar "Poodle", whereas a "Weimeraner" may elicit the exemplar "Pointer". In either case, the exemplar elicited would enable the person to properly (and naturally) categorize the animal as a dog.

Probabilistic approaches to categorization, especially exemplar theories, account for a number of experimental findings. Most importantly, they account for the robust finding that fuzzy boundaries exist between categories. Probabilistic approaches take the average typicality of a set of features and judge membership based on feature similarity. They allow concepts to be categorized based on achieving a category specific similarity threshold, regardless of the different individual features involved in reaching that threshold.

A number of models have used feature-weighing techniques to create nonlinear categories and represent probabilistic exemplars with modest success (Anderson 1990, 1995; McClelland and Rumelhart 1986). Neural networks, such as ALCOVE, do a very good job modeling exemplar categorization (Kruschke 1992). Artificial intelligence models have used brute force engineering methods such as nearest neighbor modeling to simulate exemplar categorization. In both cases, these models compare the sample object with all stored exemplars and then either use the most similar exemplar (neural networks) or the totality of similarity (artificial intelligence) to judge representativeness. These models store multiple instances of a category in memory. When presented with a new instance, artificial intelligence models would compare that information with all stored examples and build a membership threshold based on how similar the new example was to all of the stored examples. Neural networks utilize similar reasoning but use the most similar referent to make a category decision.

A number of recent examples demonstrate the power of exemplar models in technology. While Amazon.com's early attempts at a recommendation engine relied on prototype theory, their new models employ exemplars. This change in theory has provided Amazon.com with one of the most powerful commerce engines in the world. Utilizing a consumer's purchases over time, Amazon.com is able to generate multiple exemplars that lead to very precise recommendations that often result in a sale.

Google employs robust exemplar models to derive category information for its search and advertising systems. Their advertising products rely on a knowledgebase that maps and categorizes colloquial language in a hierarchical structure that employs exemplar categorization using technology principally derived from WordNet (Miller 1999) and expanded upon by Simpli.com, Inc. Google then overlays website content to the knowledgebase and derives exemplars for each webpage. Google uses those exemplars to recommend keywords, search listings and advertisements. This technology is at the heart of Google's revenue engine and is principally derived from exemplar theory.

Utilizing this category model solves two fundamental problems for Google. First, it enables Google to categorize seemingly disparate pieces of information into a best-fit category based on exemplars. Second, it enables Google to suggest alternatives that are not necessarily specific matches. An important mechanism in this technology is a hierarchical, network-based approach to knowledge organization. The technology assumes that knowledge is often arranged from general to specific. In fact, almost every word in natural language has both generalizations and specializations of this kind. These relationships form a network structure and are at the heart of language and category technology models such as WordNet (Miller 1999) and CYC (Lenat 1990).

The power of a network representation is that it puts specific information in a more general framework that can be used to compute answers to queries. For example, suppose the sample network is used to generate an advertisement. If content on a webpage activates the prototype for a Porsche and the exemplar for a 911, only the 911 node is activated. This problem is intractable for companies such as Google and Yahoo!, as there are rarely enough advertisers to support specific exemplars. Google, for example, may not have an advertisement for the Porsche 911. However, by using a network structure that can activate multiple exemplars as well as category information, i.e., the transportation, sports car and Porsche nodes, Google is able to expand its advertisement universe.

Exemplar models of categorization do a very good job with fuzzy categories. However, these models fail to account for non-perceptual features, such as function and other explanatory information. A series of models built by Wisniewski (Wisniewski and Medin 1994; Wisniewski 1995) suggest that explanatory relations play a significant role in the categorization process. This is a robust finding that includes a number of explanatory constructs: context (Medin and Schaffer 1978), prior information (Wisniewski and Medin 1994), theory (Medin, Wattenmaker, and Hampson 1987; Medin and Ortony 1989) and function (Medin and Ortony 1989; Putnam 1988; Rips 1989).

Generally, probabilistic theories cannot account for these findings. Furthermore, these theories fail to account for data that suggests that there are features that play a more central role in categorization. Consider the finding that when people are asked whether an object is a member of a category associated with its exemplar, they often say "No", if it is missing certain key elements (Malt 1994; Medin and Ortony 1989; Rips 1989; Stibel 2006). Models based on probabilistic approaches cannot account for these findings.

3. Philosophical essentialism and technology innovation

One of the most intractable problems with the aforementioned theories, as well as some of the technologies they have enabled, is the fact that people tend to only look at a single dimension when categorizing novel objects or artifacts (Medin, Wattenmaker, and Hampson 1987; Stibel, in press). The argument, first put forth by Parmenides (Stace 1928), simply asserts that nothing that is one can have plurality, as plurality requires separation and diversity. For that reason, the mind can only consider things on a single dimension, context, or function.

This has led to a long-standing debate as to whether natural kinds and artifacts have "essences" that determine category membership. This view, dubbed "essentialism" by Putnam (1988, 1990), states that an object ceases to be a member of a particular category if and only if it is determined that the object does not adhere to the same essence as do other paradigmatic members. Aristotle (1941) originally argued for essentialism in his metaphysics when he claimed that parts of a substance can only be potential substances and cannot be considered apart from the whole. Thus, in order for features to become realized, they must exist apart from the whole object. As an example, Aristotle considered whether parts of an animal could be considered as individuals. For instance, some animals live after they have been divided. However, it is clear that in such animals there are potential lives existing only as a part to the whole. Only once the animal has been divided, will the potential be realized. Thus all parts exist potentially when united in a whole.

This is one of the reasons why search engines do not exclusively rely on probabilistic models of categorization. While recommendation engines can make predictions based on similarity, search is typically keyword driven. People tend to search for essential characteristics. Search engines — including Yahoo!, Google, Lycos and Ask Jeeves — generate input in the form of a keyword from users and the engine returns a list of related results based on that keyword. That is the core principle of essentialism — a feature or set of features defines category inclusion.

Several studies, however, have demonstrated that essences do not sufficiently account for category membership (Braisby, Franks, and Hampton 1996; Malt 1994; and indirectly, Walker 1992). Psychologists (Bloom 1996; Medin and Ortony 1989; Rips 1989) developed a weaker version of the strong essentialism defined by Putnam, which better accounts for the experimental data. "They argue that people *believe* there are essences shared by sets of things, and this belief plays a role in determining what things will be accepted as members of a category and called by the category name" (Malt 1994: 66). For natural kinds (see, for instance, Medin and Ortony 1989), this essence is typically determined by essential perceptual features and physical components (what Aristotle called a "material cause"); whereas for artifacts (see Bloom 1996), essences are usually derived from function and intent of creator (Aristotle's "final" and "efficient" causes).

What this means for categorization is that the mind, if it is to consider parts (or features) of objects, must have a way in which to unite those parts into a unified whole. This philosophical postulation explains a great deal of experimental results. For instance, when people are given a set of low correlated features, they tend to focus on single dimensions when given categorization tasks (Stibel 2006). Medin et. al. (1987) showed that people not only avoid a family resemblance approach to categorization, but they also tend to ignore multiple features when making category decisions. Subjects were asked to divide sets of objects into two categories. Even though these objects varied in the degree to which their features were similar to one another based on a family resemblance model, subjects still divided the categories based on a single feature. Moreover, the features that subjects tended to choose were arbitrary; they were neither more nor less central to the object. This goes contrary to a probabilistic account. Even when Medin et. al. (1987) incorporated functions, explanations, and labels, subjects still tended toward a single dimension.

A number of technical models have been proposed to incorporate essences and explanation (Dejong and Mooney 1986; Wisniewski and Medin 1994). Essentialist models approach categorization from a much different perspective than do probabilistic models. Whereas probabilistic models hold features as central to categorization, these models assumes that an underlying essence binds categorization and features are used, peripherally, to determine whether an object has a particular essence. For this reason, essentialist models can account for a vast number of findings. For example, psychological essentialism has been shown to account for context effects (Medin and Schaffer 1978), influences of prior knowledge (Wisniewski and Medin 1994), and functional aspects (Bloom 1996; Medin and Ortony 1989; Rips 1989) that play a role in category acquisition.

Junk email or "spam" blocking software relies heavily on a hybrid model that incorporates essentialism to avoid false positives. Utilizing natural language processing, spam blocking software identifies key determinants of spam (i.e., sender, keywords, subject matter), incorporates algorithms that compare these determinants and then makes a decision based on probability. This probabilistic account works very well for spam in general because it can identify offending email based on similarity to other spam. However, clever spammers work hard at avoiding detection and trick these systems by imbedding content in their email that appears to be legitimate. To counter this approach, spam blocking software employs psychological essentialism to create fixed criteria that get at the essence of spam (i.e., number of times the email was sent or received).

Essentialism has also been well documented in child psychology and replicated in technology aimed at children. For instance, Markman (1989) showed that children tend to assume that names refer to whole objects (single units) rather than specific features. Spelke, Von Hofsten, and Kestenbaum (1989) showed that even infants use perceptual information to unite distinct features and view them as one-dimensional objects. Thus, the child is able to see the object as a whole, not a collection of parts.

LeapFrog has built an entire company around software programs aimed at children and created a model for categorization that mimics development. They utilize essentialism in most of their toys precisely because of the way children think about categories. LeapFrog's most rudimentary toys for babies employ linearly separable categories that are often derived from single features. Their Baby Counting Pal, for example, associates numbers to colors, while their StoryBlock Book teaches children to match numbers, colors, and animals to their essential characteristics. Toys for older children increase in complexity utilizing more characteristics but are still based on psychological essentialism.

While essentialism does account for certain anomalies that could not otherwise be accounted for, it fails to account for the far greater number of categories that were accounted for by probabilistic theories. Numerous experiments have shown that features play a crucial role in categorizing most novel objects. For instance, psychological essentialism cannot explain why things with high H2O content (the perceived essence of water) are or are not considered to be water (e.g., lakes versus coffee or tap water versus saliva) (Malt 1994). Furthermore, they cannot account for the fact that, cross-culturally, essences do not remain consistent (Walker 1992).

One conclusion to draw from categorization research is that features cannot be looked at as distinct from the objects to which they belong; rather, objects must be considered a group of tightly inter-connected features. This has broad implications for theories of categorization. To account for this, models must not only take into account individual features but also look at the totality of the features and essence. This entails a unified approach to modeling how the mind categorizes artifacts and natural kinds.

To address this, a number of models have simply attributed feature interconnectedness to how the features appear in nature. Early memory models took advantage of this fact (Anderson 1990; McClelland and Rumelhart 1986). These models rely on the high interconnectivity between features to account for feature similarity within a semantic hierarchy. The features, due to their interactions with one another, create a highly interconnected set. Psychologically, this accounts for why people are able to observe and utilize a family resemblance structure (Rosch and Mervis 1975).

Interconnectivity amongst features is a robust finding; in fact, perceptual features are taken to rely on each other within natural kinds (see, for instance, Medin and Ortony 1989). Normative data suggests that natural kinds have features that are highly correlated with one another (Gonnerman, Anderson, Devlin, Kemper, and Seidenberg 1997). While animals have internal parts that interact and rely on each other, this is not the case for many artifacts. Artifacts do not tend to have highly interconnected sets of features. Categorization of artifacts appears to rely heavily on the central role of functional explanations.

To account for artifacts, a number of models assume that features are united through theory, function, or prior knowledge (Wisniewski and Medin 1994). Greatly oversimplified, objects with feature sets that are not physically correlated are connected through psychological beliefs. While this approach falls short of accounting for natural kinds, it has had success in artifact categorization. Whereas the wings of a bird are naturally correlated with its other features (feathers, beak, legs, etc.), the wings of an airplane do not appear immediately connected to the rest of its features (has windows, made of metal, transports people, etc.). However, once the aspect of flying is incorporated into perceptual knowledge, the features can be coherently united to form a single object.

This approach is precisely how technology companies are moving into the next generation of software development. Hybrid models of categorization are currently being employed by Microsoft, Google, and Yahoo! to advance search engine technology. Natural Language Processing (NLP) software firms, such as Nuance, Inc., employ hybrid techniques to analyze and interpret human language. Government agencies, such as the United States' CIA and the United Kingdom's MI5, are utilizing hybrid models to categorize documents and communication and identify threats to national security. This approach codifies the various experimental results of category formation and provides adequate grounds for building models of human categorization. The mind, in viewing the world, must decipher a plethora of perceptual stimuli. It must, foremost, determine boundaries and limits. In doing so, the mind instinctively utilizes salient patterns to predict category placement. As technology continues to leverage these principles, more innovations will surely follow.

References

Anderson, J.R. 1990. *The Adaptive Character of Thought*. Hillside, NJ: Lawrence Erlbaum.

Anderson, J.A. 1995. *An Introduction to Neural Networks*. Cambridge, MA: The MIT Press.

Aristotle 1941. *The Basic Works of Aristotle*. R. McKean (ed.). New York: Random House.

Aristotle. 1963. *Categories and De Interpretatione*. J.L. Ackrill (ed.). London: Oxford University Press.

Bloom, P. 1996. "Intention, history, and artifact concepts". *Cognition* 60: 1–29.

Braisby, N., Franks, B., and Hampton, J. 1996. "Essentialism, word use, and concepts". *Cognition* 59: 247–274.

Dejong, G. and Mooney, R.J. 1986. "Explanation-based learning: An alternative view". *Machine Learning* 1: 145–176.

Collins, A.M. and Loftus, E.F. 1975. "A spreading-activation theory of semantic processing". *Psychological Review* 82: 407–428.

Fried, L.S. and Holyoak, K.J. 1984. "Induction of category distributions: A framework for classification learning". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10: 234–257.

Gonnerman, L.M., Anderson, E.S., Devlin, J.T., Kemper, D., and Seidenberg, M.S. 1997. "Double dissociation of semantic categories in Alzheimer's disease". *Brain and Language* 57: 254–279.

Keil, F.C. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: The MIT Press.

Kruschke, J.K. 1992. "ALCOVE: An exemplar based connectionist model of category learning". *Psychological Review* 99: 22–44.

Lenat, D.B., 1990. "Cyc: Toward programs with common sense". *Communications of the ACM* 33: 8.

Malt, B.C. 1994. "Water is not H2O. *Cognitive Psychology* 27: 41–70.

Markman, E.M. 1989. *Categorization and Naming in Children: Problems of Induction*. Cambridge, MA: The MIT Press.

McClelland, J.L. and Rumelhart, D.E. 1986. "A distributed model of human learning and memory". In J.L. McClelland and D.E. Rumelhart (eds), *Parallel Ddistributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models*. Cambridge, MA: The MIT Press, 170–215.

Medin, D.L. and Ortony, A. 1989. "Psychological essentialism". In S. Vosniadou and A. Ortony (eds), *Similarity and Analogical Reasoning*. New York: Cambridge University Press, 179–195.

Medin, D.L., and Ross, B.H. 1990. *Cognitive Psychology*. Orlando, FL: Harcourt Brace Jovanovich.

Medin, D.L and Shaffer, M.M. 1978. "A context theory of classification learning". *Psychological Review* 85: 207–238.

Medin, D.L., Wattenmaker, W.D., and Hampson, S.E. 1987. "Family resemblance, conceptual cohesiveness, and category construction". *Cognitive Psychology* 19: 242–279.

Miller, G.A. 1999. "On knowing a word". *Annual Review of Psychology* 50: 81–97.

Minsky, M.L and Papert, S.A. 1988. *Perceptrons*. Cambridge, MA: The MIT Press.

Nozick, R. 1981. *Philosophical Investigations*. Cambridge, MA: Harvard University Press.

Plato. 1961. *Plato: The Collected Dialogues Including the Letters*. E. Hamilton and H. Cairns (eds.). Princeton, NJ: Princeton University Press.

Putnam, H. 1988. *Representation and reality*. Cambridge, MA: The MIT Press.

Putnam, H. 1990. *Realism with a Human Face*. Cambridge, MA: The MIT Press.

Rips, L. 1989. "Similarity, typicality, and categorization". In S. Vosniadou and A. Ortony (eds), *Similarity and Analogical Reasoning*. New York: Cambridge University Press, 21–59.

Rosch, E. and Mervis, C.B. 1975. "Family resemblance: Studies in the internal structure of categories". *Cognitive Psychology* 8: 382–439.

Smith, E.E. and Medin, D.L. 1981. *Categories and Concepts.* Cambridge, MA: Harvard University Press.

Spelke, E.S., Von Hofsten, C., and Kestenbaum, R. 1989. "Object perception in infancy: Interaction of spatial and kinetic information for object boundaries". *Developmental Psychology* 25: 185–196.

Stace, W.T. 1928. *A Critical History of Greek Philosophy.* London: Macmillan & Co.

Stibel, J.M. 2005. "Increasing productivity through framing effects for interactive consumer choice". *Cognition, Technology & Work* 7: 63–68.

Stibel, J.M. 2006. "The role of explanation in categorization decisions". *International Journal of Psychology* 41: 132–144.

Walker, S.J. 1992. "Supernatural beliefs, natural kinds, and conceptual structure". *Memory and Cognition* 20: 655–662.

Wisniewski, E.J. 1995. "Prior knowledge and functionally relevant features in concept learning". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(2): 449–468.

Wisniewski, E.J., and Medin, D.L. 1994. "On the interaction of theory and data in concept learning". *Cognitive Science* 18: 221–281.

Author's Address

Jeffrey M. Stibel
Web.com, Inc.
303 Peachtree Center Avenue, 5th Floor
Atlanta, GA 30327
U.S.A.

jstibel@yahoo.com

About the author

Jeffrey M. Stibel is currently President and CEO, Web.com, Inc. (NASDAQ: WWWW). He previously served as General Manager of United Online, Inc. (NASDAQ: UNTD), owners of NetZero, Juno, and Classmates.com. Stibel also serves on the Board of Directors for Brown University's Entrepreneurship Program, Thinmail, The Search Agency, Sea Vista Development, Axon Labs, and Worldwide MediaWorks. Stibel studied business, economics and brain science at MIT and Brown University, where he was a Brain and Behavior Fellow.

Crime scene investigation as distributed cognition*

Chris Baber, Paul Smith, James Cross, John Hunter,
and Richard McMaster
The University of Birmingham

Crime scene investigation is a form of Distributed Cognition. The principal concept we explore in this paper is that of 'resource for action'. It is proposed that crime scene investigation employs four primary resources-for-action: (a.) the environment, or scene itself, which affords particular forms of search and object retrieval; (b.) the retrieved objects, which afford translation into evidence; (c.) the procedures that guide investigation, which both constrain the search activity and also provide opportunity for additional activity; (d.) the narratives that different agents within the system produce to develop explanatory models and formal accounts of the crime. For each aspect of distributed cognition, we consider developments in technology that could support activity.

Keywords: crime scene investigation, distributed cognition, narrative, technology

1. Introduction

The crime scene investigation process begins with a crime being reported, and culminates with the apprehension of the perpetrator (or the exhaustion of all lines of enquiry) and the crime being filed. The following sequence of events outlines this process:[1]

1. Crime committed, crime reported, incident created.
2. Scene investigated, evidence retrieved.
3. Offender caught or identified and charged.
4. File compiled by File preparation, receive all information on the enquiry, from the Police: Witness statements, Crime Scene notes, Forensic reports, etc.

Pragmatics & Cognition 14:2 (2006), 357–385.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

5. File sent to CPS, decision made whether to prosecute.
6. Decision made to continue, file made available to defence barrister.
7. Defence or prosecution may request additional examinations during the course of the trial.
8. All information, used and unused should be made available and declared using an "unused evidence" form (submitted with the statement).
9. CSI, Police, etc. are requested to submit a statement of actions to the court. On some occasions if counsel has questions concerning the evidence they will call the witnesses to court, along with the relevant evidence.

At each stage of the investigative process, evidence needs to be collected, recorded and disseminated to the next link in the investigative chain (Smith et al. 2004). In the UK, decisions are made concerning the admissibility of evidence by a variety of organisations, e.g., Crown Prosecution Service, the Courts, The Law Officers (i.e., Attorney and Solicitor General), The Lord Chancellor's Department, and The Home Office. Clearly, crime scene investigation is a distributed process through which different organisations become involved in the evaluation of evidence. Ultimately, it is the CPS that prepares cases for Court; to ensure that prosecutions started by the police involve the right defendants prosecuted on the right charges before the appropriate court,[2] and to put cases through "the Tests".[3] Thus, when the CPS receives a file from the police, the prosecutor will read the papers and then a decision is made whether there is enough evidence against the defendant.

2. Distributed Cognition

Cognitive activity can be performed by more than one 'agent' (humans or technological) interacting within a system. This is certainly true of crime scene investigation, in which, evidence is likely to be collected by one group of people, analysed by another group, and interpreted and presented to Court by another group. Each group specialises in some aspects of the activity, and each group is subject to scrutiny in terms of its ability to perform its specialised role. Should it be possible to question the reliability of any of these activities, then the evidence can be invalidated.

For Distributed Cognition researchers, objects can be viewed as 'resources for action', i.e., the design, appearance or functionality of specific objects or the layout of specific environments will call to mind particular actions for the person (Marti 2000; Artman and Waern 1999; Bang and Timpka 2003; Busby

2001; Garbis and Waern 1999; Hutchins 1995b; Hutchins and Klausen 1998; Nemeth 2003; Seagull et al. 2003). The actions could be physical, e.g., in the sense that some objects might 'afford' picking up, or could be cognitive, in the sense that some objects might support a particular form of information-processing, e.g., price labels on foodstuffs in a supermarket might support arithmetical calculation. From this point of view, objects can be responded to in terms of physical or cognitive actions, and their presence in the world influences the activity of the people around them. The manner in which actions are performed will, of course, depend on the appearance (or States) of the objects and the Goals of the people.

2.1 Wright et al.'s resources model

Wright et al. (1996, 2000) elaborate the Distributed Cognition notion of resources for action into a 'resources model'. This model aligns particular resource types, defined in terms of abstract information structures with an interaction strategy, to propose how changing the configuration of resources can alter a person's activity. Thus, the model has three components:

i. **Abstract information structures:** The Abstract Information Structures cover the *plans* put into effect by agents, the *goals* held by agents, the *possibilities* for performing specific actions, the *history* of previous actions, perceived *action-effect relations* and the *state* of objects during the course of interaction.
ii. **Representations:** Each Abstract Information Structure can be represented in a variety of ways. Plans could be written down as procedures or instructions (or could be learned and memorized by the agent); possibilities could be presented as a set of alternative options, such as a menu; history could be recorded as a list of previous actions; state could be described by the current action of a particular product. Thus, representations are used to either convey (through the design of the artefact) or infer (through the action of the user) one or more of the abstract information structures.
iii. **Interaction strategies:** Interaction between person and artefact is a cycle of acting and checking. The person sets out to achieve a goal and looks for information from the object that looks relevant to that goal. There are several generic interaction strategies each of which draws upon a subset of abstract information structures, e.g., Plan following (plan, history, state), Plan construction (goal, possibilities, action-effects, state), Goal matching (goal, possibilities, state), History-based elimination (goal, possibilities, history).

Table 1. Relating objects of investigation to abstract information structures

	State	Goal	History	Plan	Possibility
Environment	Visual inspection	Inspect	Similar scene	Look at scene, take photographs	Contain surfaces/ objects
Surface	Visual inspection or treatment	Search	Similar scene/ sequence of search	Procedure/ technique	Contain evidence
Object	Visual inspection or other analysis*	Collect		Collect/ Record	Contain evidence/is evidence
Sample	Other analysis	Send	Database of samples	Record	Contain evidence
Analysis	Other analysis	Result		Record	Evidence can be got
Identification	Result of analysis	Identify	Previous convictions/ Stored on database	Identify	Place person at scene or for activity
Report	Photograph/lab results/Result of analysis/ Identification	Summarise	Collect previous analysis and investigation	Collate/ Present	Complete, coherent account of a crime

*'other analysis' covers any non-visual activity, such as chemical analysis (DNA, fingerprint), or through microscopy (fibres)

The choice of interaction strategy will clearly depend on a host of contextual features. However, the point is that the strategy influences the selection of abstract information structures to which the person will attend. This implies that following one interaction strategy in preference to another would also predispose the person to ignore some of the abstract information structures. This, in turn, suggests that there is a relationship between the action followed and the resources selected, as well as the earlier assertion of a relationship between the resources selected and the action to follow. In some cases, an object-in-the-world will serve as a resource for action and offer the possibility of performing a particular action; in other cases, the course of action (as defined by an interaction strategy) will lead the person to look for particular resources in order to perform particular actions. The manner in which the possible abstract information structures are represented can vary according to the context in which an action is being performed (see Table 1).

2.2 Implications for technology

As part of the MsSAM project, we are developing a variety of technologies to support CSI. Throughout this paper, examples of these technologies will be discussed with reference to particular sections (although the relevance of the technology will probably extend to other sections). From the perspective of Distributed Cognition, it is proposed that there are four main types of resource for action within crime scene investigation: the environment itself, which affords particular forms of examination; objects within the environment, which afford interpretation as evidence; the procedures that govern crime scene investigation, which afford application to different environments; the narratives constructed during the course of crime scene investigation, which afford both hypothesis formation and explanatory models of the incident. There is an ongoing debate within the CSI community as to how much the activity of a crime scene investigator can be considered as 'search' and how much it is a retrieval activity. Leaving aside the question of how retrieval can be performed in the absence of search, the issue at stake is how much leeway should be granted to the CSI in terms of interpreting objects in terms of evidential value. One school of thought argues that the role of the CSI is to recover items in a manner that is as neutral as possible, to allow interpretation to be performed by other specialists, while another school of thought views interpretation as central to the CSI role. It is not the purpose of this paper to argue for one view or the other, but it is worth noting the current debate.

In terms of applying this work to the design of technology to support crime scene investigation, we can begin to generate some specific design guidelines. The resources for action that are employed in crime scene investigation need to be represented in a manner that can support shared activity, provide some input to formal reporting, and be adaptable to different contexts. This means that the representations are required to be, in the words of Nemeth et al. (2003: 381), "reliable, informative, efficient, clear, accurate, and malleable". Seagull et al. (2003: 1521) propose five functional requirements for technology that can be employed in Distributed Cognition systems: "(a) serve as a common referent for communication, (b) provide a communal memory tool for planning, (c) serve as catalyst for collaborative and Distributed Cognition, (d) allow parallel manipulation for multiple user-groups, and (e) allow flexible content-reconfiguration". To these requirements, we can also add the need to minimise (as far as practicable) the risk of contamination to a crime scene, and the need to minimise (as far as practicable) the interruption caused by recording of findings, e.g., through note-taking.

3. Narratives as resource-for-action: Synthesising multiple representations

The notion that 'narratives' play a role in criminal investigation is certainly not new. Crime-fiction writers, from Poe, Collins and Doyle to present day, have their protagonists relate the chain of events and association of evidence that incriminates an individual. More recently, there has been interest in the ways in which narratives are used in Court (Barry 1991; Brooks and Gerwitz 1996). An obvious example of the use of narrative lies in the manner in which eye witnesses produce accounts of incidents. It is well known that such accounts are often problematic and subject to a range of confounds, biases and distortions (Loftus 1979; Wells 1985; Wells and Olson 2003). In this paper we are not going to engage with the large literature relating to eye witness testimony. More recently, there has also been a great deal of interest in the manner in which juries makes use of narrative to 'frame' events and to make sense of evidence and testimony. Pennington and Hastie (1986, 1988, 1993; Hastie and Pennington 1995) state that jurors use 'story models' to organise information in court hearings. The proposal is that such story models both help and hinder assimilation of information by framing the available information into coherent (for the listener) accounts. This might lead to associations being drawn which are inappropriate or to links being missed.

The concept of narrative is also used to consider how evidence is presented in Court (Conley and O'Barr 1990; Engestrom 1998). In particular, the notion of 'anchored narratives' has been used to show how the manner in which information is presented, discussed and summarised during the process of a trial can be analysed according to concepts from semiotics and narrative psychology (Jackson 1988, 1990, 1995, 1996; Wagenaar et al. 1993). While such approaches are not without critical commentary, e.g., Morley (1996), they do provide a means of addressing the ways in which people attempt to make sense of evidence. This issue of 'sense-making' as the interpretation of aspects of the world to produce "meaning" that can be communicated and shared is central to the entire process of criminal investigation, from reporting a crime to recalling incidents to reporting and detailing evidence. A particular challenge for our work is the notion that sense-making occurs in discourse rather than simply in minds or representations. Sense-making is a collaborative process in which different agents contribute to an interpretation of events. What is particularly difficult to articulate is a 'discourse' that covers such a highly distributed system as CSI, in which all agents have different roles and expertise, and focus on different aspects of evidence gathering and analysis. This raises the possibility of using technology to create and encourage a 'collective intelligence'. Additionally, the notion of sense-making accords very well with the notion of 'resources for action' outlined above, i.e., people make 'sense' of the world

> ...by accepting stimuli from their environment, including others around them, and interpreting what these stimuli mean. The subsequent actions, including probing of the environment, lead to additional stimuli, which must provide meaningful affordances to grab attention, and subsequent processing. Human and non-human agents must be attuned to relevant affordances in order to interpret them, to act based on them, and to probe for additional stimuli (Nosek 2001: 219).

3.1 Narrative and argument

The procedures that govern criminal investigation require formal reports to be produced according to specified standards. Thus, an Incident Log records initial details of the crime; a Crime Report compiles all the information relating to that crime; Contemporaneous Notes are recorded by Scene of Crime Officers and these are written into a Scene of Crime Report. In contrast, many investigations draw upon eye-witness testimony. The content and structure of such reports will not be subject to any 'formal' rules, but depend upon the witnesses own account. This contrast between formal and informal narratives speaks to the contrast between narrative and argument proposed by Bruner (1992, 2002). In this concept, a narrative could be viewed as a fairly loose, informal discourse that contains sufficient information to illustrate an account, while an argument could be viewed as formal representation of objective evidence. One way of thinking about this distinction is that the argument represents the paradigmatic 'information-processing' model of cognition, in which interpretation follows a definable logic and can be seen as objective; while narrative is a form of discourse in which people create stories to explain and interpret information. As Boland and Tenkasi (1995) point out, in distributed cognition, it is likely that different agents will create narratives in different ways, i.e., that various experts might have subtle differences in their definition of aspects of the environment. This could suggest that having a 'standard' form of reporting of evidence (which, as we shall see below, is something being implemented across UK criminal investigation) without an opportunity for discussion could be problematic.

3.2 Narratives and reporting of incidents

If we assume that most police activity takes place in response to a telephone request for help, then the work of Zimmerman and colleagues is instructive (Whalen and Zimmerman1990; Zimmerman 1992). Consider the following extract of narrative from a call reporting an 'incident' to an Emergency Call Centre.

Table 2. Extract of dialogue between caller and district call handler

Dis .hh Midcity emergency
Caller .hhh Yeah uh(m) I'd like tuh: -report (0.2) something
weir:d that happen:ed abou:t (0.5) uh(m) five minutes
ago, 'n front of our apartment building?

In this extract, the Caller is seeking to engage in a dialogue to define the required response, i.e., "something weir:d" that the Caller had witnessed and which might require Police response. The Caller is providing an invitation to the Call Handler to discuss the incident and determine appropriate actions. In this example, the script is being written as the conversation progresses, so the initial steps are about determining what 'ground' is available and how it might be made common. The Caller does attempt to provide some basic 'script' information (in terms of time, i.e., "five minutes ago" and location, i.e., " 'n front of our apartment building"), but these could also be viewed as invitations to the Call Handler to respond by asking for more information, e.g., precise time and address, etc. For the Caller this might be a way of defining the appropriate narrative to account for the incident; for the Call Handler this might be a way of gathering sufficient information to determine whether a crime (or event worthy of investigation) is being reported and what sort of response is required. If the response is immediate, because of suspects on the premises or immediate threat to life, then the Call Handler might seek to establish the location and initiate a response in parallel with gathering additional information.

The Call Handler will, typically, enter some information according to the structured format of an Incident Log (usually on a computer). The information will be made available to other operators, who, in turn, will manage the dispatch and activity of Police Officers (including Scene of Crime Officers). Table 3 illustrates the sort of entry that might be made in such a computer system.[4] This extract illustrates three points of interest. First, the text in capitals represents a 'real-time' summary of the dialogue between Operator (AF01) and Caller. In order for this to happen, there will be dialogue similar to that shown in table two, and this dialogue will have established a 'common-ground' which

Table 3. Extract from Incident Log

1108	8169	*AF01*	THEY HAVE PERSONS ON PREMISES WHO ARE RESPONSIBLE FOR BREAKING INTO LOCKERS. THEY ARE THERE NOW.
1109	8169	*AF01*	THERE ARE 2 MALES RESPONSIBLE THEY ARE WALKING AROUND THE FITNESS CLUB CALLER CAN POINT THEM OUT
1111	1583	*HQ00*	*XX75* Assigned (1111) Arrived (1115) Released (1125) Cancelled ( )
1111	8169	*AF01*	CALLER WILL BE AT GYM RECEPTION AND CAN IDENTIFY THE MALES. HE HAS NO DETAILED DESCRIPTION AT PRESENT

can be entered into the computer. This real-time entry means that there may well be duplication of information, e.g., the repetition of phrases emphasising that the suspects are on the premises which could require an immediate response ('persons on premises', 'they are there now', 'they are walking around the fitness club'), and the inclusion of unsubstantiated information, e.g., 'who are responsible for breaking into lockers', 'there are 2 males responsible' (these entries do not assume proof of responsibility but indicate the Caller's suspicions and the fact that there is a crime to investigate). Second, the entries in capitals form a record that can be shared by other operators, who can assign units to the incident. For example, operator HQ00 assigns unit XX75 to respond to the scene. Communication between operators and units is typically through radio communications and involves another set of dialogues. Third, the entire log forms evidence of the investigation that could be called upon in Court, and so needs to be a complete account of the activity (hence, the need to enter information in real-time).

The log records information from callers as well as the activity of the three units who had been dispatched to the site: one unit interviews a 'Male' leaving the fitness club [THE MALE GOT INTO A VEHICLE OUTSIDE THE BARRIER AND MADE OFF — UNKNOWN DESC OF CAR]; one unit checks the building [NO SIGHTING AT THE MAIN BUILDING]; and one unit finds that the suspects [LEFT VIA FRONT ENTRANCE APPRX 40 MINS AGO], together with the names, addresses and previous history of the suspects (which we need not consider in this paper). Following this activity, a decision is made to request scene of crime support [CAN WE PRINT FOR SOCO — TO PRINT THE LOCKERS #5 AS BEEN TAKEN]. Subsequently a Crime Summary Report is printed, as shown in Figure 1.

Crime incident number	Offence details	Beat	Location	Times/day/ date	Investigating officer	State of crime

Figure 1. Crime Summary Report headings

It is apparent from Figure 1 that very little of the information collected during the activity in the Incident Log is transferred to this report. We have a description of the offence and its location that could be useful to the SOCO, but little else. The Incident Log might be shared between people who are working on the response, but this does not mean that it supports collaboration. However, it does imply that current practice tends to place people as 'gate-keepers' for information, which, in turn, encourages redundancy of data entry. For example, much of the content of the incident log will be entered in the personal issue log books of the officers involved.

3.3 Sharing narrative

The notion of 'sharing' narrative between individuals, such as Call Handler and Caller discussed above, might not simply be a matter of establishing and maintaining common ground. Rather discourse raises and develops new material, with the aim of determining what might need to be shared. In terms of Distributed Cognition, the Log provides a resource for action for determining general classes of information. However, it is also apparent that the role of this Log is in the immediate performance of activity, rather than an archive to be passed on to other people. At present, sharing is real-time by voice or via a shared terminal. There is little use of graphical capability for sharing or sharing over longer periods of time or between people not immediately involved in the response, such as CSI. When information is shared, it is often the 'fixed' elements, such as location, time, names, rather than interpretative elements, such as action or motive, but could include Contentious details, such as ID of a suspect.

3.4 Implications for technology

In terms of our work, there are two broad approaches to the sharing of 'narrative' that we consider. The first simply focuses on the manner in which the Incident Log could be made available in more detail to the investigator, the second relates to capturing the investigation of a crime scene. Using details from the incident log, and updating it to the various personnel in an investigation is made easier with the introduction of TETRA into emergency service operations. In effect, this represents a broad-band communications channel, over which text, images and voice can potentially be circulated. More interestingly, this channel can be set-up to provide 'talk-groups'. This means that some information can be shared within a defined group of people. From this implementation it could be possible to allow 'communities of knowing' (to use Boland and Tenkasi's (1995) phrase) to share task specific information. For example, from the discussion following Table 3 (above), if it was known that the suspects had left the building, then it might have been unnecessary to question anyone else as to their whereabouts.

During an initial inspection of a scene, the CSI could produce an audio commentary on a video. If this was available to colleagues away from the scene, such as the Crime Scene Manager, then discussion on the form of search and recovery could take place. For example, the CSM might advise on checking only Points of Entry or might advise on recovering specific items. In this manner, the CSI and CSM would have a form of video-conference that could support discussion of activity. The video could also be used as the basis for briefing other people in the investigation team. While the notion of video-conference is commonplace, the idea of situating such communication at the scene, i.e., via wearable computers we are developing, is less well established.

4. The crime scene as a resource for action: How environment influences search

An initial step in assessing a crime scene can be crudely described as the recognition of objects as having evidential value. In other words, an environment contains a multitude of things that may or may not be relevant to the investigation. For crime scene investigation it is not feasible to evaluate everything for its potential to offer evidence, e.g., due to time constraints, condition of environment or objects, possible association of objects to activity. Consequently, there is a need to conduct some initial classification and screening of the scene and to focus the attention on objects of potential evidential value.

4.1 Narrative in the collection of evidence

The purpose of examining a crime scene is to formulate a hypothesis based on all the available evidence, therefore understanding the most likely course of events that has resulted in the observed circumstance (Jamieson 2004). As the search of the crime scene progresses, Schraagen and Leijenhorst (2001) suggest that forensic investigators use narrative to determine how best to collect evidence. The narrative might include expectations concerning evidence and expectations concerning the crime. In their study (in which trainee forensic investigators work at a simulated crime scene), they showed that 'experts' use this narrative to select appropriate heuristics to guide their search for, and interpretation of, evidence. The process by which items are selected can be considered analogous to directed search. This implies that search involves not only seeing something but also developing an expectation that something will be present.

To some extent, one could consider the crime scene as 'affording' (in Gibson's (1950) sense) items for investigation. The environment itself, therefore, becomes a resource for action in that it supports the search for evidence, not merely as a repository of things to be found but, more importantly, as an associated set of objects for investigation. The crime scene would have points of access, it would have opportunities for movement, it would have locations at which fibres or other samples might be captured, and it would have various objects that might have been handled during the course of the crime. Each of these would afford a particular response from the CSI and searching for each of these might be influenced by the expectation of particular affordances. Thus, an experienced CSI is likely to 'see' a suitable surface for fingerprints, rather than merely a table or windowsill or a good place to look for fibre samples rather than a splintered door-jamb.

The internal model of the CSI would relate objects according to a particular interpretation of the crime, the environment and other factors, such as weather conditions. Thus, CSIs might make more of a conscientious search for footwear marks according to the weather. In the Autumn/Winter months, when the ground is wet and muddy or covered in snow, footwear marks will be more readily apparent than in the Summer months when the ground might be dry, hard and resistant to impression. Similarly, the expectation of finding fingerprint marks increases as glove-wearing becomes less comfortable during the Summer or might also increase when small objects or containers that are difficult to open are found. An experienced CSI investigating the scene of a break-in may well walk around outside the building to determine the likely point of entry (POE) and routes to and from the building. This then guides the search to specific locations. Once inside the building, the CSI will look for obvious signs of a route from POE through the building, e.g., as indicated by the direction of footwear marks on the floor, or whether footwear marks are on furniture, windowsills etc., or whether objects have been moved or displaced.

In terms of the abstract information structures and interaction strategies of Wright et al. (1996, 2000), one can consider the environment to support the action shown in table one. Furthermore, the manner in which the environment is represented, for later discussion or examination, can also be considered. Photographs, sketches and brief notes are ways in which the CSI can off-load the burden of remembering details of the scene by recording them for later use. The process of creating notes, photographs and sketches also supports a particular course of action, viz. a top-down approach to the examination in which a global view is taken prior to homing in on specific items of interest.

4.2 Implications for technology

For the consideration of an environment, it might be useful to have some knowledge of previous investigations, e.g., of that type or in that area. Traditionally, this would be the province of the experience gained through years of practice, possibly in one region. However, if one assumes that not all CSIs will have similar experience, that shift in patterns, area of coverage and mobility might mean that sharing of information is less easy than in the past, then having some shared repository of previous crimes could be useful.

By logging activity and evidence against a location, it could be possible to support to collaborative activity over time. For example, we take a photograph of a scene, track an investigator's movement around the scene (e.g., through Global Positioning System or local positioning) and superimpose this track onto the photograph. This is illustrated by Figure 2. The annotation can be supplemented by additional material relating to the activity, e.g., recover item, or analysis (Baber et al. 2005a, b; Cross et al. 2004).

The annotated photographs will serve as a record of activity. However, at a deeper level, they could function as 'boundary object's (Star 1988; Boland and Tenkasi 1995), i.e., representations that provide a common frame of reference for discussion. This frame of reference could, for example, be used to guide discussion about procedure, e.g., why did you do X?, or to interpret activity, e.g., where was item Y recovered? Of particular interest to our work is the relative significance of graphical annotation in contrast with textual annotation. We feel that greater benefit accrues from graphical annotation than text, which is supported by the work of Guibert et al. (2005).

Whilst the use of annotated photographs provides information on a specific scene, it does not easily allow comparison across scenes. An obvious approach to this problem would be to create a database. However, we have decided against this for several reasons, the principal being the need to have some

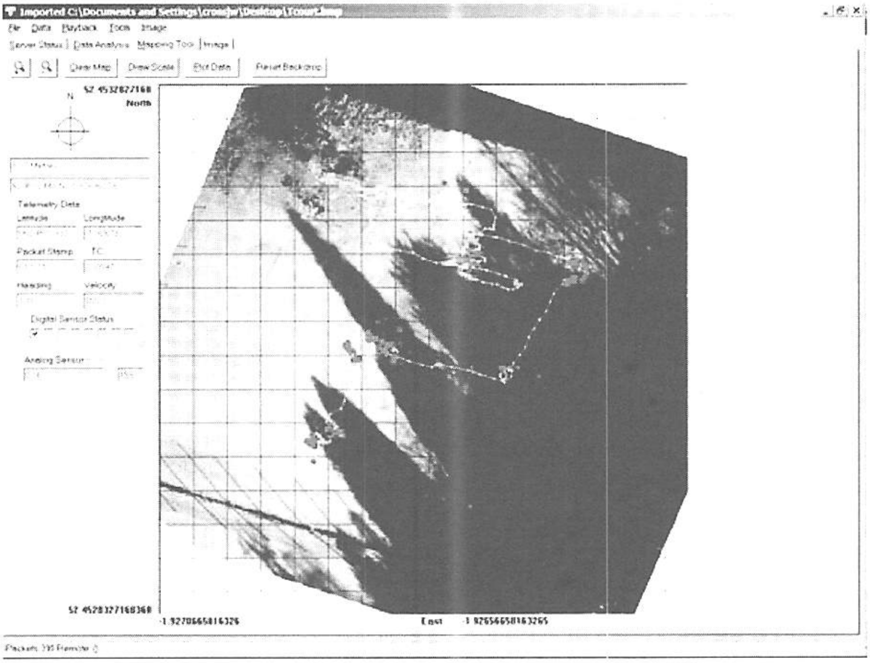

Imported C:\Documents and Settings\crossjw\Desktop\Tower.bmp
File Data Playback Tools Image
Server Status Data Analysis Mapping Tool Image
Clear Map
Draw Scale
Plot Data
Reset Backdrop
N
52.4532827168
North
Telemetry Data
Latitude
Longitude
Packet Stamp
TC
Heading
Velocity
Digital Sensor Status
Analog Sensor
52.4528327168368
-1.9270665816326
East
-1.92656658163265

Figure 2. Using position and activity of investigators to annotate a photograph. A photograph (taken from the roof of an adjacent building) forms the background for the plot of position (light dots) and item recoverry (dark dots) performed by an individual.

notion of data structure to construct the database and requiring querying of the data. Our intention is to have a means of storing information in as free a form as possible, while allowing as automated retrieval as possible. To this end, we employ Case-Based Reasoning (CBR). For example, an entry contains an address and method of entry. This is submitted, and matching performed which reveals that the same method of entry has been used on properties with similar post-codes. This might suggest a related collection of burglaries using the same modus operandi.

5. Objects as resources for action: The affordance of evidence

A central tenet of Distributed Cognition is the notion that the representations that are manipulated and exchanged by the agents within a system, make up the system's 'mental state' (Hollan et al. 2002; Flor and Hutchins 1991). A significant aspect of crime scene investigation lies in recognising the objects in the scene as potential external representations. Rather than having the information displayed to the investigator (as is the case with many of the technologies considered by distributed cognition research), the investigator needs to perceive the objects as representing some useful information. Thus, a footprint becomes useful if it can be lifted from the scene and if it can be interpreted in the context of a particular activity at the scene. Whereas much of the work on Distributed Cognition considered earlier in the paper focussed on the manner in which artefacts were designed and used to represent information, this is not the case for crime scene investigation. The objects for investigation are not given, but need to be discovered and translated. Having conducted a search, and determined the relevance of specific objects, the next step is the examination of these objects. In this step, one can consider that the objects afford (in Norman's (1988) sense) examination. In other words, the appearance of an object, coupled with the experiences of previous similar objects, affords certain activities. These activities might range from simply collecting the object and storing it for later display, to treating the object in order to allow collection for later examination in the laboratory.

In addition to objects themselves serving as representations, the investigator will also be making representations, e.g., in the form of notes, photographs, reports. For the purposes of this paper, we propose a simple dichotomy between those external representations that are used to support activity-in-the-world, which we term 'informal', and those that are mandated by procedures or organisational requirements, which we term 'formal'. During the investigation, the CSI will complete various documents. Table four shows some of the six important categories of documentation that are considered applicable to any search, in terms of their possible narrative type (formal or informal).

The UK rules relating to disclosure of evidence from a crime scene means that any records made by an investigator must be made available to both defence and prosecution counsel in court.[5] Disclosure entails maintenance of carefully logged records reflecting specific time points in a scene investigation and in relation to other actions. It also means that recording will be made in a manner that reflects as much certainty as possible, with as much supporting evidence as available. If there is doubt or uncertainty, then the evidence could easily be called into question. *All* material collected at a crime scene ought to be disclosable. However, this does not mean that all material *will* be disclosed (but it must be available, if so required), e.g., not all of the photographs collected during crime scene investigation will be submitted as evidence. The reason for this might be technical, e.g., poor image quality, or redundancy, e.g., several versions of the same view, or quantity, e.g., don't submit too many images, or

Table 4. Documentation as Narrative Types

Artefact	Content	Narrative type
Administrative Worksheet	Major events in investigation	Formal
Description of scene	General appearance of scene	Informal
Photographic log	Record of photographs taken	Formal
Diagram/Sketch	Drawing of scene	Informal
Evidence Recovery log	Record of objects and evidence retrieved from the scene	Formal
Latent print lift log	Record of prints lifted	Formal

relevance, e.g., only submit images that support the report being made. Thus, photographs collected at the scene are 'informal' until they have been subjected to a selection process and then the selected photographs become 'formal'. The photographs must retain their integrity, i.e., not be tampered with if they are to be subject as evidence. This means that the photograph itself does not change, but that its status as an external representation could move from informal to formal. By a similar token, the content of notes and sketches made at the scene remain ostensibly the same, but are copied into a formal version in a crime scene report.

From a broad perspective, one can suggest that the investigation proceeds through initial examination of the environment, on to the selection of appropriate objects to examine and the collection of particular samples. The samples are subject to analysis and the result reported and used in the identification of an individual. In this section, we consider a specific type of evidence (fingermarks) in terms of their role as resources for action.

5.1 Fingermarks

In order for a fingermark that may or may not be visible to become represented as a fingerprint, it needs to be discovered. This can be done by dusting (brushing powder over), or performing some chemical treatment of surfaces that might carry a fingerprint. The selection of an appropriate surface becomes a matter of judgement at the scene; one would not expect to scan all surfaces. Rather, the selection of likely surfaces to look for fingerprints must be performed. Even when a site has been selected, finding a fingerprint is not necessarily straightforward. For example, there might be smeared marks, there might be several prints that overlap etc. At this juncture, the CSI must decide where there are likely to be latent fingerprints, which can be subjected to testing. Some surfaces are not conducive to conventional fingerprinting. The CSI would evaluate whether items or surfaces are suitable for fingerprinting with powders, e.g., non-porous, clean surfaces, and whether to use flake powders which are less destructive to the surface and surrounding materials and easier to clean, or granular powders e.g. on outside drainpipes, or with chemical treatment, e.g., paper. If the print is visible to the naked eye, it can be photographed as part of a sequential process of recovery, and with the increasing digital imaging capabilities the marks may be captured and sent electronically. However, it is highly likely that the CSI will wish to have a physical lift as well as the digital image.

The process of 'discovering' a fingermark is not simply a matter of looking with the naked eye. As mentioned above, the surface requires some form of treatment. Observations of CSI suggests that there is often a painstaking cycle of examination, treatment and checking. For example, a bright light shone obliquely across the surface to help visually scan the area for prints. Used in conjunction with the powdering techniques, the light helps reveal contours and ridges in the mark. When using flake powders the process of dusting typically requires the gradual building up of layers, e.g., using small circular motions of the brush, until a satisfactory impression is produced. This is inspected again under the light and if necessary dusted further, until it is ready to be lifted. Lifting involves placing clear sticky tape over the mark and removing. There is research into digitising the process of lifting, i.e., having a camera capture the image of the mark (although it is a moot point as to whether this will ever remove the need to preserve the mark itself through traditional methods).

Once a print has been preserved and recorded, it can be analysed. Analysis might be performed on site or (more likely) at a laboratory remote from the site. The analyst can benefit from knowing some background information pertaining to the print, e.g., the location and orientation of the print. If the print is made by someone wearing gloves, it might be possible to collect a print if the pattern on the material looks sufficiently distinctive — if the same pattern recurs then this might be sufficient to place a person at a particular set of locations.

One can imagine a continuum over which fingerprints might be considered. At one extreme, a fingerprint identifies an individual; at the other extreme, a fingerprint is a mark on a surface that may or may not be visible to the unaided eye. The transition across this continuum, of course, defines the process of crime scene investigation, i.e., the movement from evidence to identification. However, it also highlights some relevant aspects of Distributed Cognition, namely the manner in which phenomena are represented and these representations shared across a system.

1. Person: affords production of latent print;
2. Action: affords making latent print;
3. Surface: affords latent print being held;
4. Powder/chemical: affords latent print being visible/developed print/mark;
5. Print/photograph: affords collection of developed print;
6. Analysis: affords pairing developed print/mark with known individual and affords reducing uncertainty of pairing developed print/mark with person.

The process of collecting a fingerprint, therefore, culminates in pairing the person who made the print (1.) with a known person with an acceptable degree of confidence (6.). In order for this process to be effective, it is necessary to make certain assumptions about the production of fingerprints, e.g., in terms of where, in the environment, one might expect to find such evidence (3.) and what action might have led a person to produce fingerprints in such a location (2.). Having made these assumptions, the CSI seeks to capture and collect the developed print (4. and 5.). From the point of view of Distributed Cognition, each representation could be held by different agents or objects at the scene or during different phases of investigation. The point at issue is how well the crime scene investigation procedures allow sharing and transfer of representations. The selection of a specific 'good' print (at stages 1. to 3.) could have a bearing on stage 5., but it is not always communicated. There might be useful information to be gained from knowing that a print was made on the underside of a window rather than on the ledge or on the glass itself. The transition from 2. to 3. and from 3. to 4., might introduce artefacts into the analysis. Table 5 presents the fingerprint in terms of abstract information structures and representation.

The notion that search involves an internal model was discussed in the section relating to the crime scene as a resource for action. It was proposed that the purpose of the internal model was to provide the investigator with a means by which items could be looked for or the environment scanned. The assumption is that, rather than recovery of evidence being entirely based on the 'objective' search of a scene, there is likely to be some influence of expectation and experience involved in the search. A good CSI will probably recognise these effects and manage them. However, the influence of experience and expectation could even creep into areas of analysis which are usually supposed to be entirely 'scientific', i.e., free from bias. For example, in a recent study of fingerprint analysis, Dror et al. (2005) explored the influence of contextual factors on the matching of pairs of fingerprints. In this study, contextual factors were manipulated by the story and photographs that were used to explain the

Table 5. Relating Fingerprints to Abstract Information Structures and Forms of Representation

State	Goal	History	Plan	Possibility	Representation
Visual inspection of surface either before or after treatment	Find, capture and collect latent prints	Recall previous encounters with similar scenes	Look at scene, use powder or chemicals to reveal print, capture prints, log, pass on to laboratory	Environment contains surfaces and objects that might yield fingerprints amenable to lifting	Surfaces can hold prints, Reveal latent prints, Lifted print held on medium and labelled, Log, Analysis, Interpretation

source of the fingerprints, e.g., crimes with no physical harm to the person versus crimes with extreme physical harm. The study showed that in cases where the fingerprints were unambiguously different, there was little effect of context. When the fingerprints were ambiguous, i.e., when the certainty as to whether they were the same of different decreased, then the contextual factors seemed to play a role in increasing the likelihood of seeing a match. However, this effect was only observed for the context in which extreme physical harm featured in the background story. The study suggests that in cases where there might be some uncertainty as to whether fingerprints match and where the crime is extreme, that matching might be influenced by context.

5.2 Implications for technology

The recording of the environment and collection of objects becomes the main input to a computer system for the CSI. As a photograph is taken or an object placed in an evidence bag, the activity is recorded in terms of date/time and position. Position is supplied either through Global Positioning System (GPS) or through a bespoke positioning system (Cross 2004). The evidence bag itself is tagged with a Radio Frequency Identification (RFID) tag that can be read when it is picked up and data written to (see figure three). Thus, when a CSI picks up an item and places it in an evidence bag, a report is already being written in terms of CSI, date, time, place, type of evidence. The CSI can then add additional information, perhaps in the form of a verbal description to the evidence. Furthermore, by providing at least the crime report, the CSI would have the basic information required to find the scene and begin examination. It is a moot point as to whether any additional incident or activity narrative would be beneficial (or whether this would induce bias in the collection of evidence).

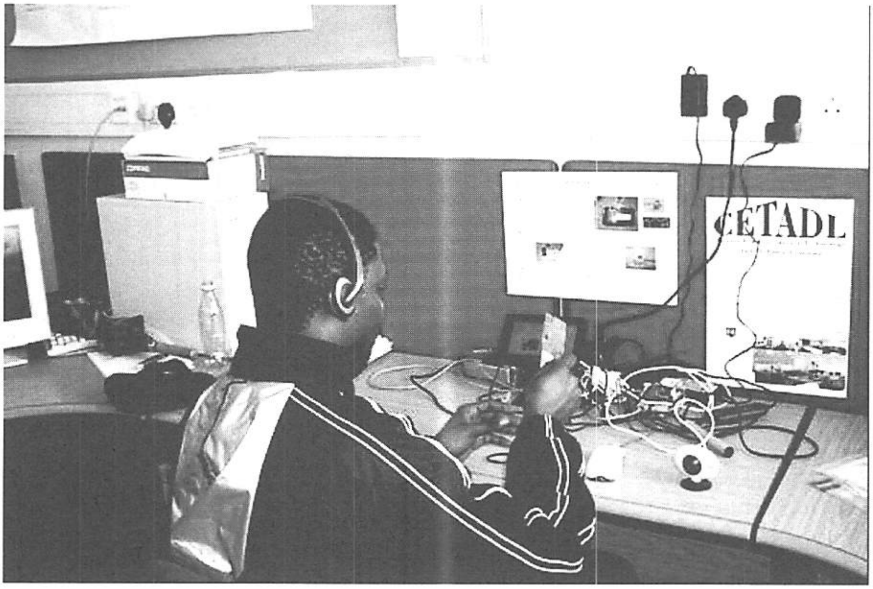


Figure 3. Interacting with objects tagged with RFID, using speech to input descriptions

6. Procedures as resources for action: Distributing crime scene investigation across people, procedures, and technology

Crime scene investigation has evolved in accordance with procedural changes. These changes have been implemented, amongst other reasons, as a consequence of circumstances uncovered during the course of an investigation, as a result of technological advancement, or as a corollary of the increased awareness of contamination issues. A successful outcome to forensic investigation is dependent on the structured and systematic appraisal of the scene, reliant on the practitioner identifying, selecting, recovering and submitting the evidence for testing, in accord with accepted procedures and providing a clear documented record of the examination (Jamieson 2004). To this end, there is a need for the forensic practitioner to be aware of system limitations, and how the product to system amalgamation has to be modified to meet the needs of a socially oriented judicial system (Williams 2004).

While much of Distributed Cognition has concentrated on the use of artefacts as resources for action, Suchman (1987) pointed out that plans and procedures could be treated in the same way as artefacts, i.e., as resources to be interpreted and applied, rather than programs to be followed. This introduces a potential point of tension for crime scene investigation: procedures may be written with the intention of providing strict guidance on the manner in which evidence is to be collected and crime scenes examined. However, the enormous variation in crime scenes (not to mention pressures placed upon CSIs from different types of investigation) mean that the procedures may not be universally applicable. Thus, a CSI might view procedures as inherently flexible, because of the need to adapt a given procedure to the unique aspects of each crime scene. However, if procedures are written to be followed to the letter and, as such, intended to be programs or rules to be adhered to, then it is possible to question performance when 'off-procedure', which could result in evidence being rejected. Thus, there might be a tension between procedure as program and procedure as resource. Indeed, the role of the CSI has been described as being ambiguous (Wright 2002), and often the direction is torn between the practical demands and official procedural requirements (Hobbs 1988). There are occasions when procedures and investigative parameters stipulated by governing bodies can conflict with the environmental constraints, and operating practicalities of the investigative process.

The point is that it is possible for procedures to function as both prescribed step-by-step instructions and as resources for action. Whenever procedures are presented as step-by-step instructions, there is an attendant possibility that violation might arise, and given the adversarial nature of British Law, this might lead to the authenticity of a particular piece of evidence being called into question, i.e., if it can be shown that the CSI deviated from procedures, then the collection of evidence could be questioned. On the other hand, the procedures might not be suitable to all situations and so require some adaptation (which might be construed as violation if the procedures were written in a very prescriptive manner). A possible solution to the problem would be to present procedures as goals, e.g., ensure recording or avoid contamination, etc. In this way, the account would not be how well the steps were followed but whether goals adhered to and how. However, the end result would be to 'Collect valid evidence from the crime scene', which, in itself requires the CSI to adhere to principles of 'best practice' regarding contamination of evidence, without the need for instructions prescribing the order in which actions are carried out.

This discussion returns to the earlier consideration of narrative versus argument. The aim of the investigation is to provide sufficient evidence, collected in a rigorous and objective manner, to help develop an argument. In this view, crime scene investigation is a process of induction, i.e., from specific evidence to an explanation. The procedures, and the constraints they place on CSIs, can be seen to be designed to follow this inductive process. However, this is in contrast to the fact that much of the investigative process is deductive, i.e., using hypotheses to determine what information might be beneficial or useful. As we saw in the discussion of objects-as-resources-for-action, a CSI might select objects in terms of their potential to be (or provide) evidence. This decision-making could be seen to be somewhat tangential to the aims of the procedures. If a CSI followed the procedures to the letter, their activity could be readily seen in the light of Bruner's notion of argument, whereas much of the exploration would appear to follow the notion of narrative.

6.1 Implications for technology

To some extent having the means to explain and comment upon activity, such as the audio commentary on a video, could be seen as a useful way of providing narrative within the framework of argument. In this way, the CSI would be able to explain choices made at the scene. While this might satisfy the theoretical concerns raised in the previous section, it does lead to the practical concerns of disclosure raised earlier. If the CSI deliberated over a course of action or piece of evidence, this could be taken to indicate uncertainty — which in Court could be used to indicate a lack of professionalism.

7. Discussion

The suggestion that narratives can be passed through the crime investigation processes suggests a degree of collaboration (see Table 6). As Olson and Olson (2000) point out, technology to support collaboration requires (amongst other aspects) the ability to support common-ground, coupling and to encourage collaboration readiness. Within the Computer-Supported Cooperative Work (CSCW) community, the notion of Common Ground draws upon the work of Clark (1996). This views common ground, between two speakers, as "…the sum of their mutual, common or joint knowledge" (p. 93). In any discourse, it is important for speaker and listener to establish sufficient overlap of knowledge to enable them to refer to a 'common ground'. The speakers do not spend time working out what is common but assume it to be so (unless the assumption is disproved and then some repair is needed). This process could draw upon cultural knowledge to make assumptions about beliefs, actions, language etc. of individuals, e.g., a Caller might attempt to speak in 'police jargon' such as the use of ICAO alphabet for registration plates; or could be personal, in that we share points of perceptual reference, e.g., when the speaker says something like '*that* night', both parties know what is being referred to.

Table 6. Representation of information in the Crime Scene Investigation System

	State: environment	State: object	State: perpetrator	State: victim
Incident	Verbal (caller) Textual (999 Ops)	Verbal (caller) Textual (999 Ops, caller)		Verbal (caller) Textual (999 Ops)
Activity	Verbal (caller) Textual (999 Ops, Attending officer)	Verbal (caller) Textual (999 Ops, Attending officer)		
Crime	Textual (999 Ops, Attending officer)	Textual (999 Ops, Attending officer)		Textual (999 Ops, Attending officer)
CSI	Graphical (CSI) Photographic (CSI) Textual (CSI) Numerical (CSI)	Graphical (CSI) Photographic (CSI) Textual (CSI) Numerical (CSI)	Graphical (CSI) Photographic (CSI) Textual (CSI) Numerical (CSI)	Graphical (CSI) Photographic (CSI) Textual (CSI) Numerical (CSI)
Forensic	Graphical (Forensic) Photographic (Forensic) Textual (Forensic) Numerical (Forensic)	Graphical (Forensic) Photographic (Forensic) Textual (Forensic) Numerical (Forensic)	Graphical (Forensic) Photographic (Forensic) Textual (Forensic) Numerical (Forensic)	Graphical (Forensic) Photographic (Forensic) Textual (Forensic) Numerical (Forensic)
Report	Textual (Attending officer, CSI, Forensic	Textual (Attending officer, CSI, Forensic	Textual (Attending officer, CSI, Forensic	Textual (Attending officer, CSI, Forensic
Case	Verbal (caller, Barrister, attending officer, CSI, forensic)	Verbal (caller, Barrister, attending officer, CSI, forensic)	Verbal (caller, Barrister, attending officer, CSI, forensic)	Verbal (caller, Barrister, attending officer, CSI, forensic)
Accused	Textual (accused, witnesses) Verbal (accused, witnesses)	Textual (accused, witnesses) Verbal (accused, witnesses)	Textual (accused, witnesses) Verbal (accused, witnesses)	Textual (accused, witnesses) Verbal (accused, witnesses)

From this view, a challenge CSCW is to provide a means by which people can negotiate (or otherwise work with) 'common ground'. Unfortunately, the notion of common-ground is somewhat problematic when applied to crime scene investigation: on the one hand, it implies a mutual understanding of all relevant material (which is likely to be desirable); but on the other hand, it implies a common view of the situation (which is likely to be undesirable in that it suggests a degree of biased interpretation). The process of collecting evidence is supposed to be performed independently, so that when all the evidence is gathered together, it can be interpreted objectively. Furthermore, the very nature of crime scene investigation means that the entire process is geared towards constructing 'common ground' (or a case that can be tried in Court). The key issue would appear to be how best to define the elements which are essential to share.

The technologies developed during this project aim at supporting collaboration throughout the CSI process. At the scene itself, collection of evidence is used to automatically log activity and record exhibit details. This automatic log is made available to colleagues, to support discourse related to ongoing activity, but can also be re-used by investigators later, i.e., by having all information stamped with location data. We view such developments as the logical extension of the increasing digitisation of CSI and the increasing communications bandwidth offered by TETRA (and similar broad-band radio systems). By considering the narrative, collaboration and resources-for-action, we believe that it is possible to engage in theoretically-driven, user-centred design of future CSI technology.

Notes

* The work reported in this paper is supported by a grant from the UK Engineering and Physical Science Research Council (EPSRC — GR/S85115: MsSAM Methods to Support Shared Analysis for Mobile Investigators).

1. The process outlined in this section, and considered throughout the paper, is taken from UK practice. There may well be differences in some aspects across different countries and we are not claiming superiority for UK practice, merely acknowledging this is the source of information used in our work.

2. In England and Wales, all cases start off at a magistrate's court, and the magistrate takes the decision as to whether a 'higher' court is necessary, according to the offence). It is also possible for the accused to opt to have their case tried in a higher court.

3. **The evidential test** — the CPS must be satisfied that there is enough evidence to provide a "realistic prospect of conviction". If the evidence is not strong, the CPS asks: Can the evidence be used in court? Is the evidence reliable?; **The public interest test** — There are instances when looking at the public interest will lead to a decision not to prosecute.

4. As far as possible the format and content of the tables in this paper are drawn from incidents handled by a UK Constabulary. Under the constraints and restrictions of the Data Protection Act, it is not possible to exactly reproduce the material and so modifications have been made. Items in italics have been altered, but do not substantively alter the content of the table for the purposes of the discussion.

5. Section 3 (1) of the Main Provisions of Criminal Procedure and Investigations Act, 1996 clearly states that, "*The prosecutor must — (a) disclose to the accused any prosecution material which has not previously been disclosed to the accused and which in the prosecutor's opinion might undermine the case for the prosecution against the accused, or (b) give to the accused a written statement that there is no material of a description mentioned in paragraph (a)*".

References

Artman, H. and Waern, Y. 1999. "Distributed cognition in an emergency co-ordination center". *Cognition, Technology & Work* 1: 237–246.

Baber, C., Cross, J., and Houghton, R.J. 2005. "Distributed collaboration in a remote search task: Comparison of datalink and audio communications". In P.D. Bust and P.T. McCabe (eds), *Contemporary Ergonomics 2005*. London: Taylor and Francis, 64–68.

Baber, C., Cross, J., and Smith, P. 2005. "Methods to support shared analysis for mobile investigators". In *International Workshop on Annotation for Collaboration*. Paris: Centre National de la Recherche Scientifique, 11–20.

Bang, M. and Timpka, T. 2003. "Cognitive tools in medical teamwork: The spatial arrangement of patient records". *Methods of Information in Medicine* 42: 331–336.

Barry, A.K. 1991. "Narrative style and witness testimony". *Journal of Narrative and Life History* 1: 281–293.

Boland, R.J. and Tenkasi, R.V. 1995 "Perspective making and perspective taking in communities of knowing". *Organization Science* 6(4): 350–373.

Brannigan, E. 1992. *Narrative Comprehension and Film*. London: Routledge.

Brooks, P and Gerwitz, P.D. (eds). 1996. *Law's Stories: Narrative and Rhetoric in the Law*. New Haven, CT: Yale University Press.

Bruner, J. 1992. "Psychology, morality, and the law". In D.N. Robinson (ed), *Social Discourse and Moral Judgment*. San Diego, CA: Academic Press, 99–112.

Bruner, J. 2002. *Making Stories: Law, Literature, Life*. New York: Farrar, Straus, and Giroux.

Busby, J.S. 2001. "Practices in design concept selection as distributed cognition". *Cognition, Technology & Work* 3:140–149.

Clark, H.H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Cohen, L.J. 1977. *The Probable and the Provable*. Oxford: Clarendon Press.

Conley, J.M. and O'Barr, W.M. 1990. *Rules versus Relationships: The Ethnography of Legal Discourse*. Chicago, IL: The University of Chicago Press.

Cross, J., Baber, C., and Woolley, S.I. 2003. "Layered annotations of digital images for data collection in the field". In *Digest of Papers of the 7th International Symposium on Wearable Computing*, Los Alamitos, CA: IEEE Computer Society, 154–159.

Dror, I.E., Péron, A., Hind, S-L., and Charlton, D. 2005. "When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints". *Applied Cognitive Psychology* 30: 799–809.

Engeström Y. 1998. "The tensions of judging: Handling cases of driving under the influence of alcohol in Finland and California". In D. Middleton and Y. Engeström (eds), *Cognition and Communication at Work*. Cambridge: Cambridge University Press, 199–232.

Flor, N.V. and Hutchins, E.L. 1991. "Analyzing distributed cognition in software teams: A case study of team programming during perfective software maintenance". In J. Koenemann-Belliveau, T.G. Moher, and S.P. Robertson (eds), *Empirical Studies of Programmers: Fourth Workshop*. Norwood, NJ: Ablex, 36–64.

Garbis, C. and Waern, Y. 1999. "Team coordination and communication in a rescue command staff: The role of public representations". *Le Travail Humain* 62: 273–291.

Gibson, J.J. 1950. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Grusenmeyer, C. and Trognon, A. 1997. "Les mécanismes coopératifs en jeu dans les communications de travail: Un cadre méthodologique". *Le Travail Humain* 60: 5–31.

Guibert, S., Darses, F. and Buijut, J-F. 2005. "Using annotation in engineering design: Some results from an experimental study". *International Workshop on Annotation for Collaboration*, Paris: Centre National de la Recherche Scientifique, 61–72.

Hastie, R. and Pennington, N. 1995. "Cognitive approaches to judgement and decision making". *Psychology of Learning and Motivation* 32: 1–31.

Hollan, J., Hutchins, E., and Kirsch, D. 2002. "Distributed cognition: Toward a new foundation for human–computer interaction". In J. Carroll (ed), *Human–Computer Interaction in the New Millennium*. New York: Addison-Wesley, 75–94.

Hutchins, E. 1995. "How a cockpit remembers its speed". *Cognitive Science* 19: 265–288.

Hutchins, E. and Klausen, T. 1998. "Distributed cognition in an airline cockpit". In Y. Engeström and D. Middleton (eds), *Cognition and Communication at Work*. Cambridge: Cambridge University Press, 15–34.

Jackson, B.S. 1988. *Law, Fact, and Narrative Coherence*. Merseyside: Deborah Charles.

Jackson, B.S. 1990. "Narrative theories and legal discourse". In C. Nash (ed), *Narrative in Culture*. London: Routledge, 23–50.

Jackson, B.S. 1995. *Making Sense in Law: Linguistic, Psychological, and Semiotic Perspectives*. Liverpool: Deborah Charles.

Jackson, B.S. 1996. *Making Sense in Jurisprudence*. Liverpool: Deborah Charles.

Jamieson, J. 2004. "A rational approach to the principles and practice of crime scene investigation: Principles". *Science and Justice* 44: 3–7.

Loftus, E.F. 1979. *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.

Morley, I.E. 1996. "Narratives, anchored narratives and the interface between law and psychology: A commentary on Jackson (1996)". *Legal & Criminological Psychology* 1: 271–286.

Nemeth, C. 2003. "How cognitive artefacts support distributed cognition in acute care. In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA, 381–385.

Nosek, J. 2001. "Social construction of knowledge in teams: Issues for computer support". In M. McNeese, E. Salas, and M. Endsley (eds), *New Trends in Cooperative Activities:*

Understanding System Dynamics in Complex Environments. Santa Monica, CA: Human Factors and Ergonomics Society, 218–229.

Olson, G. and Olson, J. 2000. "Distance matters". *Human Computer Interaction* 15: (2–3) 139–178.

Pennington, N. and Hastie, R. 1986 "Evidence evaluation in complex decision making". *Journal of Personality and Social Psychology* 51: 242–258.

Pennington, N. and Hastie, R. 1988. "Explanation-based decision making: Effects of memory structure on judgment". *Journal of Experimental Psychology: Learning, Memory and Cognition* 14: 521–553.

Pennington, N. and Hastie, R. 1993. "The story model for juror decision making". In R. Hastie (ed), *Inside the Jury: The Psychology of Juror Decision Making.* Cambridge: Cambridge University Press, 192–221.

Schraagen, J.M. and Leijenhorst, H. 2001. "Searching for evidence: Knowledge and search strategies used by forensic scientists". In E. Salas and G. Klein (eds), *Linking Expertise and Naturalistic Decision Making.* Mahwah, NJ: Lawrence Earlbaum, 263–274.

Seagull, F.J., Plasters, C., Xiao, Y., and Mackenzie, C.F. 2003. "Collaborative management of complex coordination systems: Operating room schedule coordination". In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society,* Santa Monica, CA, 1521–1525.

Smith, P., Baber, C., Wooley, S., Cross, J., and Hunter, J. 2004. "A task analysis of crime scene investigation". In *Contemporary Ergonomics 2004.* London: Taylor and Francis, 331–335.

Star, S.L. 1988. "The structure of ill-structured solutions: Heterogeneous problem-solving, boundary objects and distributed artificial intelligence". In M. Huhns and L. Gasser (eds), *Distributed Artificial Intelligence, volume 2.* Menlo Park, CA: Morgan Kaufman, 37–54.

Suchman, L.A. 1987. *Plans and Situated Action.* Cambridge: Cambridge University Press.

Wagenaar, W.A., van Koppen, P.J., and Crombag, H.F.M. 1993. *Anchored Narratives: The Psychology of Criminal Evidence.* London: Harvester Wheatsheaf.

Wells, G.L. 1985. "The eyewitness". In S.M. Kassin and L.S. Wrightsman (eds), *The Psychology of Evidence and Trial Procedure.* Beverly Hills, CA: Sage, 43–66.

Wells, G.L. and Olson, E.A. 2003. "Eyewitness testimony". *Annual Review of Psychology* 54: 277–295.

Williams, R. 2004. "The management of crime scene investigation in relation to the investigation of burglary and vehicle crime". *Home Office Findings 235.* London: Crown Copyright.

Wright, A., 2002. *Policing: An Introduction to Concepts and Practice.* Devon: Willan.

Wright, P.C., Fields, B., and Harrison, M.D. 1996. "Distributed information resources: A new approach to interaction modelling". In T. Green, J. Canas, and C. Warren (eds), *Ecce-8: Proceedings of the Eighth European Conference on Cognitive Ergonomics,* Le Chesnay, 5–10.

Wright, P.C., Fields, R.E., and Harrison, M.D. 2000. "Analyzing human–computer interaction as distributed cognition: The resources model". *Human–Computer Interaction* 15: 1–41.

Authors' addresses

Chris Baber
The University of Birmingham
Birmingham
United Kingdom

Paul Smith
The University of Birmingham
Birmingham
United Kingdom

James Cross
The University of Birmingham
Birmingham
United Kingdom

John Hunter
The University of Birmingham
Birmingham
United Kingdom

Richard McMaster
The University of Birmingham
Birmingham
United Kingdom

About the authors

Chris Baber has been a lecturer at the University of Birmingham since 1990 and was promoted to Reader in 2004. He is the author of some 50 journal papers and over 200 conference papers or book chapters. His research concerns human interaction with technology, in particular through the use of everyday skills and abilities. This work includes the study of speech-based interaction with computers and wearable computing, and has been funded by the UK Ministry of Defence, EPSRC and the European Union. He is a member of the EPSRC College, and has been on the Scientific Committee of the IEEE International Symposium of Wearable Computers.

Paul Smith has been a Crime Scene Investigator with Leicestershire Constabulary since 1999. He received an MSc in Forensic Archaeology from the University of Bournemouth in 2004. A Registered Forensic Practitioner (RFP) and a qualified Crime Scene Manager and Coordinator, he is currently seconded to the University of Birmingham to assist in developing methods to support shared analysis for the forensic investigator. He is also working towards a PhD in Wearable Technological Applications for Crime Scene Investigation.

James Cross graduated from the University of Birmingham with an MEng in Electrical Engineering, and went on to obtain a PhD in 2003. His main research interests lie in the development of wearable computing, imaging systems and hardware design.

John Hunter was appointed Professor of Archaeology and Ancient History at the University of Birmingham in 1996. Apart from following an extensive scheme of research excavation and survey in Scotland, he began to develop forensic archaeology in 1988, co-wrote the first textbook on the subject and has routinely lectured to police and forensic groups ever since. He founded the Forensic Search Advisory Group, is a Director of the Centre for International Forensic Assistance (CIFA), a lead assessor for the Council for the Registration of Forensic Practitioners (CRFP), and member of CRFP's Incident Investigation panel. He works actively on police and war crime missions throughout the world and has recently returned from the Balkans as a consultant to the International Commission on Missing Persons (ICMP).

Richard McMaster received a BSc (Hons) in Psychology and then an MSc in Occupational Psychology from the University of Sheffield in 1999 and 2002. He joined the University of Birmingham in 2004 as a Research Associate and is also undertaking a part-time PhD in Distributed Cognition in the Emergency Services.

Web search engines and distributed assessment systems*

Christophe Heintz
Institut Jean Nicod, Ecole de Hautes Etudes en Sciences Sociales

I analyse the impact of search engines on our cognitive and epistemic practices. For that purpose, I describe the processes of assessment of documents on the Web as relying on distributed cognition. Search engines together with Web users, are distributed assessment systems whose task is to enable efficient allocation of cognitive resources of those who use search engines. Specifying the cognitive function of search engines within these distributed assessment systems allows interpreting anew the changes that have been caused by search engine technologies. I describe search engines as implementing reputation systems and point out the similarities with other reputation systems. I thus call attention to the continuity in the distributed cognitive processes that determine the allocation of cognitive resources for information gathering from others.

Keywords: web, epistemology, trust, reputation, epistemic authority

1. Introduction

Hutchins' (1995) seminal example of distributed cognitive system is the system implemented in the organisation of a navy ship, where numerous individuals and artefacts have their own specific cognitive tasks that contribute to the overall goal of the system: directing the ship. The distributed cognitive systems I analyse in this paper are made of the community of Web users, who link documents, and a search engine. The overall task of the systems is to enable an efficient use of the cognitive resources of Web users. These systems, I argue, are distributed assessment systems. In these cognitive systems, the cognitive task of Web users (as authors of Web pages) is to assess Web documents, and the cognitive task of the search engine is to compile these assessments and produce a usable representation of the result.

Pragmatics & Cognition 14:2 (2006), 387–409.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

In the first section, I introduce the problems that one wanting to acquire information from others must face. The cognitive function of distributed assessment systems is then described as providing solutions to these problems.

In the second section, I argue that search engines are reputation systems: they attribute to web documents 'reputation labels' in the form of ranking on given queries (Search Engine Result Pages, SERPs). Search engines provide visibility to web documents, and this visibility directly determines the trusting behaviour of search engine users. An essential aspect of second generation search engines is that their ranking algorithms take as input the linking behaviour of web-users. The consequence is that search engines together with web-users constitute a distributed cognitive system for the attribution of reputation, visibility, and, eventually, credibility.

In the last section, I compare search engines' procedures for the attribution of credibility with the procedures of scientific institutions for the implementation of meritocracies and conclude that they are essentially similar: both rely in the same way on numerous judgements from the individuals so as to compile a representation of the cognitive worth (in degree and range) of sources of information; both constitute the institutional frame of a distributed cognitive system that organises information. In the end, reputation labels from different reputation systems are compiled by end users. As a consequence, reputation systems influence each others' ranking. This contributes to reinforce the continuity between pre- and post-search engines eras. I thus question anew the extent to which search technologies of the web change the distributed cognition that informs decisions in information search. I conclude with general considerations on the relations between cognition and technology.

2. Information seeking behaviour and cognition — the advent of the Web

2.1 Consequences of the growth of online documents

The number of documents that can be found on the web is huge and there is moreover a phenomenal growth of the number of web sites, which amounted to 217% between 1998 and 2002. Today, Google claims to be indexing 8 billion pages. This phenomenon is much more than just a change in the medium of communication: producing and publishing Web documents is becoming ever easier,[1] with the consequence that the published documents actually come from very different sources of information. In particular, publishing on the Web is not restricted to recognised experts but is also open to lay people. Access to online information is, in principle, not restricted to experts either, and is open to people that may lack the discriminative abilities that would lead them to read or trust only those documents that are worth it. These two facts have caused some intellectuals to worry about the growth of the Web and its usage, especially emphasizing the need to filter the huge amount of information that is so easily accessible (e.g., Eco 2003). The normative and political aspects of questions related to the development of the Web seemed to have caused some dramatisation and overemphasis of the changes that the technologies of the Web have induced or could provoke.[2] So one goal of this paper is to re-balance a possibly distorted view and emphasise the continuity between pre-Web and post-Web epistemic practices.

The techniques of the Web include two major innovations other than electronic publishing: the hyperlink and search engines. The hyperlink (or simply 'link') enables the web users to immediately go, through a simple click, to the linked document. It thus amounts to a reference, which at the same time provides simple access to the refereed document. This technology leads to the practice known as web surfing: a specific practice for the search of information. Also, links among Web documents give to the entire Web a structure whereby documents are characterised not by their content, but by their position in the structure, i.e., by which documents link to them and which documents they link to. The structure of the web is promptly exploited by search engines that use it to get information about the documents.

Search engines (Google type) are interactive web sites that respond to queries — in the form of sets of key-words — with an ordered list of URLs: the Search Engine Result Pages (SERP). Their role has incredibly increased in the last years. It has increased economically, as is made obvious by the current battle among the biggest software companies for promoting their own search engines, and the enormous sums of money involved. But also, and firstly, its role in epistemic practices has become, within a growing range of population, major, and is still becoming more and more important. By epistemic practices, I refer to the routinised behaviours that aim at acquiring or producing new information. There is no doubt that searching for information through the use of search engines constitutes a widespread epistemic practice, as is manifest by the number of queries received by search engines. For instance, at the beginning of 2001, Google answered more than 100 million searches per day. Thus, there is a major change in epistemic practices, and its origin is the implementation of new technologies for managing information.

The success and technical developments of search engines can be understood as a consequence of the growth of information on the Web, including the following facts:

- the quantity of Web documents lies largely beyond what an individual user could ever consult;
- users could not access a Web document without knowing its URL or being automatically directed to it;
- users require some direction in deciding which websites to consult for solving their specific problem.

The reliance on search engines by individuals increased with the number of Web documents, and the number of individuals relying on the Web for the acquisition of information has also increased, hence the importance to analyse and assess the epistemic impact of search engines.

2.2 The stake: Acquiring information from others and cognitive economy

Because there are so many sources of information, the problem arises as to which source one should devote one's scarce cognitive resources. The problem can be framed in economic terms: people searching for information will have to spend time and effort for acquiring the information. Acquiring information from others is often a very good, cost-effective, way to acquire information. Think about asking your way in the street instead of trying all the streets around and making your own map of the area. Attending to and understanding others may appear to have a negligible cost with regard to the cost of acquiring the information by oneself. Yet, this cost is much increased when one factors in the risk of being deceived. Imagine you were close to the street you wanted to go to, but some passer-by misled you and sent you in the opposite direction. In this case, trusting somebody else causes you an important waste of time and effort. Also, the cost of asking and attending to some answer is rendered much more salient and worthy of interest in 'competitive' situations. Imagine you are in a street with many passers-by. Among them are some who know the area and could inform you well, but some others know nothing about it, and some others wrongly think they know and might mislead you. If you ask to somebody who doesn't know, you will not be misled, but you may miss the opportunity to ask somebody who knows. And you must also be careful not to ask someone that will mislead you. But how can you know in advance which passer-by is a good informant? You may have a few strategies such as 'do not ask someone who looks like a tourist, unless he holds a map in his hand', or 'avoid those who seem to be in a hurry'.

Users of the Web are in a similar situation: they intend to acquire information from others via the online documents, but need to pick up and select documents among all the accessible ones. There are numerous strategies for searching information on the Web and selecting online documents. One such strategy is to start with an already known document, and browse through the hyperlinked documents by clicking on those hyperlinks that seem to answer one's question best. Another strategy is to search outside the Web — e.g., asking a friend — for information about where to find specific information in the Web. Often, diverse strategies are used in concert, depending on the goals and beliefs of the information seeker. However, a strategy which is widely used consists in using a search engine. It is a strategy for information seeking that relies on new technologies, and which is the focus of this paper.

Strategies for acquiring knowledge from others range from total gullibility to radical scepticism toward what others say. The strategies are to be understood as a way out of the dilemma that people gathering information from others must face. The dilemma is as follows: Before attending to others' communications, one wants to be sure that they are worth it, i.e., that one will derive sufficient satisfaction from the time and effort invested in processing the communication (listening or reading, understanding, and drawing the relevant consequences). But how can we know in advance whether a source of information shall provide valuable information? The obvious way to know whether the information is valuable or not, is to assess it. But assessing the information requires spending some cognitive effort that we wanted to spare in the first place by relying on what others say. Likewise, when you wonder whether to buy a yogurt in a shop you may ask first: Is the yogurt worth its price? But you may have no other solution than to pay for it so that you can try it and judge by yourself. So you need to pay the price of the yogurt to know whether it is worth it, and whether you were right to buy it. For information consumption, the problem is even more dramatic, because the assessment of the value of a piece of information does not come with its consumption, as in the case of the yogurt. For instance, attending to and understanding someone's communication does not necessarily reveal whether the information communicated is true. The communicator can lie or have false beliefs. So it seems we need to deploy other means so as to check the truth and general worth of what is communicated, and these means can be as costly as finding the information by oneself. The dilemma is therefore: either one consumes the information without knowing its worth, and runs the risk of being completely misled, or one checks the information communicated, and thus renounces all the advantages of drawing information from others. Attending to others' communication seems either hazardous or useless.

As already suggested, the way out of the dilemma consists in checking the likeliness that the source of information shall provide worthy information, rather than checking the information itself. The value of the source of information guarantees the value of the information it communicates. This 'method' still includes the risk of spending even more effort checking the source of information than the information itself. For instance, a reductionist standpoint in social epistemology would require knowledge acquired from others to be rationally justified with knowledge acquired by oneself. This may be attained with a full and costly check of the reliability of the source of information, but one soon realises that information about sources of information is itself often acquired from others, and the checking appears to be a never-ending process. Reductionists thus seem to advocate a non-effective process of knowledge acquisition from others. The unfeasibility of a reductionist checking is even more apparent in cases where numerous sources of information are available: should one go through a reductionist check of all the available sources so as to choose the most reliable one?

In actual life, people use strategies that do not comply with the radical criteria of the reductionist. The stake of these strategies is not the attainment of the completely justified certainty that reductionist philosophers would aim at. Rather, the stake is to find information with sufficient guarantees that the information is sufficiently worthwhile for one's purpose, but at a cognitive cost that is minimal.

2.3 Epistemic practices on the Web

Information search on the Web epitomizes the above problems. The Web puts its users in front of a large amount of information out of which they have to discriminate and select worthwhile information. The best strategy for answering Web users needs and to maximize the utility gained with the information acquired, is certainly not to browse every Web document so as to eventually select the best ones.

Web users need to rely on information about the information published on the Web, so as to adeptly direct their cognitive effort. However, one sees that the solution can lead to an infinite regress, since one may need also to have information about the information about the information and so on. The information about information must therefore be readily exploitable without further need to go to a meta-level.

What are the actual epistemic practices of Web users — their ways to gather information and acquire knowledge? As mentioned above, Web users heavily rely on search engines. Search engines, indeed, provide information about Web documents that is directly exploitable for deciding about the allocation of cognitive resources.

Using a search engine is a strategy that is cost-effective (note that this does not mean that it is a strategy that is satisfactory in all contexts). In particular, using a search engine is a strategy of information search that includes a reliability check of the sources of information. This is due to the fact that search engines' algorithms factor in an assessment of the reliability of the sources (see Section 3 on PageRank). At this point, a key epistemological question arises: is the reliability check that goes with using a search engine satisfactory? In other words, are search engines reliable sources of information about the reliability of the sources of information of the Web? In order to give an answer, one first needs to analyse what information is given by search engines, and how this information is produced.

2.4 The cognitive function of search engines

I have presented above the difficulties related to the cognitive and epistemic task of acquiring information from others. The problem consists in choosing the source of information to which one will devote further cognitive resources. The best solution to the problem, i.e., the more satisfactory choice for a person, is the one that directs cognitive resources to information that actually has cognitive value for that person. I have shown that an essential and challenging part of the task consists in finding cost effective ways to acquire worthwhile information from others. Search engine technologies aim at facilitating, or even enabling, a cost effective information acquisition. They arose (first generation, starting in 1990, and second generation, 1997) at those points in the development of the Web, where the critical mass of information available would make it too costly to search for specific information. In economic terms, this amounts to saying that search engines answered a demand for effective ways to search information on the Web. In cognitive terms, one can say that search engines fulfil a cognitive function, which was needed for the achievement of a cognitive task of growing importance — searching for worthwhile information on the Web as a means to attract visitors and consequent economic benefits. According to theories of situated cognition (see, e.g., Clark 2003), the cognitive processes of information search on the Web are distributed between the individual and the search engine: both process representations and transmit information to each other via the key board and the screen. The eventual finding of worthy information is the product of their conjoined processes. Cognitive phenomena that determine the allocation of cognitive resources in people's quest for information on the Web thus span much further than the head of the individual seeker of information.

The goal of my account of the problem related to the acquisition of information from others was to specify the cognitive function of search engines as elements of information seeking cognition. Because of the limited resources one has, in terms of time and cognitive power to process information, one cannot deal with all the possibly worthwhile information available. I have presented the use of search engines as one of the strategies of information seekers on the Web, which satisfactorily reduce the risk of wasting one's cognitive resources in misleading sources of the information. In the next section, I analyse the cognitive part of information seeking processes that is outside the head of the individual information seeker. Search engines can be simply conceptualised as cognitive artefacts used by individuals for information search. But distributed cognition extends farther than the human–computer system: it is the whole community of Web authors that is put at work.

3. The Web's assessment systems

3.1 Search engines as meta-information providers

The information given by search engines about Web documents takes the form of lists of URLs as answers to queries, which are strings of terms keyed in a text box field. In order to understand the information provided by search engines, one can picture them as enormous registers, where entries include all terms that appear in the Web, sets (i.e., unordered) and lists (ordered) of terms. An ordered list of all the documents of the Web is associated with each entry. In actual cases, some restrictions need to be added to the initial metaphor: for instance, it is the URLs rather than the Web documents that are listed (except for the option 'cache'). Also, the URLs listed are only those of indexed Web documents, thus ignoring the mass of documents of the 'hidden Web'. Another restriction is that the algorithms of search engines are such that they will not provide an ordered list of *all* the URLs of the indexed sites, but stop at, say, the thousandth URL. It is also worth noting that the above information is not stored as such, but the couples (entry, list of URLs) are computed 'on the spot', as responses to queries, which are specified calls on entries.

The information contained in these enormous registers is made readily exploitable thanks to the following interactive interface: one keys-in one's query, and the search engines return the list of the first ten (or more) URLs of the associated entry, as well as the possibility to view the rest of the list of URLs. The hierarchy of URLs makes it easy to consult the first answers and ignore the other ones. It thus promptly reduces the set of alternatives to consider thus making decisions on information search processes that are based on manageable data.

With these ordered lists of URLs, search engines distribute ranks in query-indexed SERPs. Ranks provide users with meta-information about the information of Web documents. High ranks are explicitly presented, by Google for instance, as signs of quality and relevance with respect to given queries. Search engines thus provide reputation to Web documents in the same way as academic institutions distribute titles. Academic titles function as guarantees of competence in specific fields, while SERPs' ranks function as guarantees of cognitive worth with regard to specific topics. A SERP rank is tantamount to reputation because it encourages Web users to have a favourable behaviour towards the Web documents having the high ranks. High ranked documents will be given, on the whole, more cognitive attention than low ranked ones. This means that the ranking functions as indicator of cognitive worth. SERPs give an idea of the worth of a document before the users actually assess the document itself, and ranks are labels of the reputation of Websites. Using search engines is therefore assessing *a priori* the cognitive worth of the Web documents.

There is also some ambiguity in search engines' ranking regarding the inferences and processes underlying assessment of information and their sources. Clicking on a URL sends us to a Web document, and what is actually ranked by search engines are URLs rather than documents. But URLs cannot be equated with documents, since the same URL may host different and evolving documents. The extreme case is when no document at all is stored at a URL, while the search engines had calculated its ranking on the basis of a previously existing document. Everybody using a search engine has at least once been frustrated by such results. What is ranked, therefore, is not the Web document itself, but its source as instantiated by the URL. URLs are owned by the people publishing on the Web, so it can be assumed that the source of the documents posted at a given URL, the publisher, is stable across the different documents that are published at this URL. This is what justifies search engines to rank URLs on the basis of the documents that are stored there. The inference step is classic in assessment processes: one judges the producer, or the source of information, on the basis of what it has produced or the information it has delivered. And the converse inference is also thought to be warranted: if the source of information has proved to provide good information, then it will continue to do so. Of course, beliefs about sources of information need be constantly updated; and so search engines constantly 'crawl' the web to update their information on documents published and their sources of information. This updating and the relative perennial endurance of web documents even allow search engines to present their SERPs as lists of documents rather than lists of sources of information. The link to the 'cache' — which is a copy of the document that was assessed by the search engine and which is stored on the local machine of the search engine — gives a further justification to the inferential step: for instance, if a link is broken, the cache explains why the URL figures in the list in spite of the fact that it provides no information. Remark also that for the rapidly changing documents that are blogs, another procedure is implemented: blog search engines manage information about updating and new posting through 'ping technologies'.

3.2 Distributed cognition on the Web

The ranking in SERPs, in order to be informative, must reveal something about the documents ranked. This must be truly useful information about the information contained in the Web documents. The algorithms of search engines do not only peer into each document of the Web in order to abstract information about it. They also use the linking behaviour of Web users. Here is a brief idea of what the algorithms of the Google search engine do. Firstly, it calculates on the basis of the occurrences of terms: it is assumed that the more a document contains occurrences of a given term, the more this document is cognitively worthy for the user who specified her interest with the same term in a query. The great innovation of the so-called second generation search engines, however, is to derive information about a document not only from the document itself, but also from the documents that link to it. Thus the famous PageRank algorithm gives a quality mark calculated on the number of links to the document and the quality mark of these documents. The idea that makes the PageRank algorithm a calculator of quality is that every link from one document to another can be interpreted as a positive vote from the linking document for its target. In Google's words:

> PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important". Important, high-quality sites receive a higher PageRank, which Google remembers each time it conducts a search.[3]

Note that the interpretation of linking behaviour as positive voting remains controversial. Also, Google itself alternates between 'having value', 'importance' or 'quality' as the best interpretation of 'being linked to'. Let us say, then, that PageRank's output is a judgement of the cognitive worth of documents — as independent of their topic and the interest of the Search engine user. But the main point is that search engines do not provide such judgements through their own evaluations of Web documents; the judgements express the ones of the Web users who publish documents with hyperlinks.

With respect to relevance — the fit between the search engine user interests and the topic of Web documents — algorithms take the occurrences of the terms in Web documents as input. But the judgement of Web users is, again, also exploited: the algorithm Hypertext-Matching Analysis takes as input the content of 'neighbouring' documents so as to situate it into thematic clusters of hyperlinked documents.[4]

Because search engines' algorithms take as input the links among Web documents, they rely on the linking behaviours of Web users. These behaviours stem from the users' assessments of the appropriateness of linking to specific documents. Search engines therefore compute query-indexed ranks for Web documents on the basis of all these cognitive processes of individuals assessing Web documents and linking.

Search engines, together with the hyperlinking webpage authors, constitute cognitive distributed systems. Search engines are faced with a cognitive task: producing a representation of the cognitive worth of Web documents. In order to complete this task, they impart the sub-tasks of assessing Web documents to the many Web users that have browsed through the Web, made their own judgment, and created hyperlinks accordingly. The cognition that provides the ranks in SERPs is thus distributed among the search engine algorithm and the many people that create hyperlinks. The distributed cognitive systems — search engines + users — produce and compile multiple assessments: they are distributed assessment systems.

3.3 What the Web's distributed assessment systems do for us

The beneficial functions of the Web's distributed assessment systems are cognitive and epistemic. This sends us back to the first section of the paper. Recall that decisions in information search face what appears to be a dilemma: either checking all the information by oneself, measuring the comparative values of pieces of information and eventually select the most worthwhile — this process, when possible at all, requires an enormous amount of cognitive effort that renders trusting others worthless and repetitive; or select some information without checking beforehand its content — but then what guarantee is there that the selected information is sufficiently good?

Web distributed assessment systems provide the setting for implementing a strategy that escapes the dilemma. The rankings of Web documents are representations of their cognitive worth that are readily exploitable. They can, and they actually do, guide the decisions of information search. Because representations of cognitive worth are closely associated with availability, these representations are readily exploitable and need not call on an third level of information about the information the rankings bear. In other words, what is at work in the exploitation of representations of cognitive worth is not an understanding or deciphering of the content of these representations; what happens, rather, is that availability directly determines users' distribution of cognitive effort without necessarily involving the meaning of the availability. Search engine users, indeed, have usually little knowledge of the processes that underlie ranking. The warrant of the cognitive worth of a document is not necessarily represented as such in the head of the information consumer, but it is present in the environment as a determinant of behaviour. The cognitive processes of search engine users provide an instance of bounded rationality, which relies on a simple heuristics with quickly halting procedures rather than complying with the demoniac rationality of some philosophers and economists (Gigerenzer et al. 1999). Because humans are cognitively limited, they cannot access and comprehend all the information that is available to them. But they seem to rely on processes that manage their limited cognitive resources (Sperber 2005; Gigerenzer et al. 1999). In particular, humans (as other animals) give cognitive attention to what is at hand, what is promptly present and most salient in their environment. Search engine interfaces exploit this 'lazy' management of information search: Facing SERPs, users will click on the first available links, those that are at the top of the hierarchy and that received the best rankings. Other documents are given much less attention, if any. The most common strategy consists in browsing the Web documents beginning with the one at the top of the hierarchy and going down the hierarchy, reading the title and excerpts from documents, clicking the most promising ones, and stopping the search as soon as one gets a relatively satisfactory document.

What makes, however, the behaviour of the users of search engines rational? To what extent do these users rationally manage their cognitive resources in their search of information? I hold that Web distributed assessment systems actually enable maximizing the cognitive effect in information search. The reason for this relies on the notion of meritocracy.

First, the act of ranking constitutes the institutional setting where rewards and penalties amount to going up and down the hierarchies, acquiring better or worse ranks. This furnishes a motive for publishing worthy documents. Publishing is an act that (generally) aims at communicating with a large audience. It is rightfully assumed that the URLs at the top of the hierarchies have more chance to get clicked, and the associated Web page visited, than those at the bottom. So the chances of getting to the targeted audience are increased by good ranking in search engines' SERPs. If publishers of the Web aim at being read and carefully estimate the role of search engines, then they will want to have high rank. But high rank, as explained, is obtained through individuals' assessment of cognitive worth. This supposedly ensures that only worthy documents will be linked to. Eventually, publishers should aim at publishing worthwhile documents if they want to be read. The motivation to get a high rank in SERPs is not to be under-evaluated. It generated, for instance, a new job specialisation in websites promotion, which includes ranking optimisation. There are many discussions and arguments about how to increase 'artificially' one's web site rank — where 'artificially' means 'without receiving hyperlink references that result from the positive assessment by Web users". It appears also that, at present, commercial web sites owners are much more worried about ranking than other sources of information. Future developments of this situation remains unknown. I would like however to note that the Web remains a huge source of non commercial information on which people increasingly rely. With this increased reliance among different communities, it is likely that ranking will become motivating among these different communities. For instance, while the scientific community is nowadays much preoccupied with impact factor and quotation rate, the intrusion of the Internet and associated search technologies (see, e.g., GoogleScholar and ISI Web of Knowledge) promises to raise the awareness of the importance of rankings. Additionally, the possibility of cheating and other unsatisfactory ranking are a challenge to be met by search engines that have the imperative goal to satisfy users, if only because of fierce competitions. Attracting users is necessary because search engines attract business clients by showing that they are popular. These facts present the Web as a competitive arena where meritocratic reward systems are being implemented. The merit, I have argued, is measured in terms of cognitive worth, and it is what anyone participating in the competition should aim at. Of course, this description of the factors which contribute to the implementation of an ideal meritocratic ranking should not blind us to its malfunctions.

The real stake, eventually, is not just the motivation that authors may have to publish worthy documents, but the actual implementation of a meritocratic system. With search engines, the meritocratic ideal takes the form of an isomorphism between SERPs returned on queries and cognitive worth relativised to the topic expressed by the queries. Admittedly, even this ideal is not perfect, since cognitive worth is a value that depends more on the users' interests as defined by their situation and background history rather than on a disembodied topic of interest. It is yet another challenge of search engines software to take this fact into account, since queries can by themselves express only this abstract aspect of individuals' interests (and ways to personalise searches are being investigated). Yet, there is also a sense in which cognitive worth should include a universal measure: Inasmuch as truth and quality is not dependent on individuals,[5] so cognitive worth has an element that does not depend on people. This universalising idea is also present in the concept of meritocracy: merit does not depend on the person assessing it. Likewise, multiple distributed assessment counts on the fact that what is worthwhile for someone is also worthwhile for the others, and that, consequently for the design of search engines, a hyperlink made by one person is useful information for others. This assumption seems reasonable — all the more so given that the hyperlinks taken into considerations for queries' answers are the ones of people contemporary to the search engine users and with interests in topics related to their queries. Moreover, the objectivity of the judgement output of search engines is ensured by taking into consideration the judgement of numerous agents — the subjective biases are supposed to cancel each others.

The quality of the judgement, on the other hand, is ensured by giving more weight, in the final decision, to sources of information (which are associated with a URL) that have already been recognised as worthwhile, and so, trustworthy. Ideally, people would devote cognitive resources to the documents that have the most cognitive worth. This, indeed, would result in a maximisation of cognitive resources, since an investment in cognitive effort would be maximally compensated by the value obtained from processing the document (reading, understanding, memorizing, etc.). Search engine's SERPs are intended as a means to achieve such maximisation: meritocratic ordering of cognitive worth leads the users to give cognitive effort to the most worthy sources of information.

Are the new technologies of communication implemented by the Web truly bringing changes in the cognitive processes of information search and selection? In particular, do Web distributed assessment systems implement revolutionary processes? Often, debates on the advantages and drawbacks of the advent of the Web dramatise the changes that it brings with it. For instance, Umberto Eco (2003) reacts strongly against the epistemic anarchy that, according to him, characterizes the Web. He longs for the use of authoritative filters.

Roberto Casati (2003), on the other hand, praises the dynamics of the Web and the voting systems of search engines that initiate feedback on the values of Web documents. However, describing in more abstract terms the principles at work in information selection on the Web shows that there is much continuity and similarity between pre-Web publishing and post-Web publishing.

4. General properties of distributed assessment systems and the specificity of search engines

4.1 Other reputation systems

Here is the characterisation of the Web distributed assessment systems we have arrived at: The systems produce representations, which are assessments of the cognitive worth of Web documents. The production involves not only the algorithmic computations of the search engines, but also the mental computations of those who surf the web and produce Web documents with hyperlinks. We hold that scientific institutions have developed similar complex processes that seek to optimize both quality and relevance of the sources of information to which scientists dedicate time and effort.

The individual scientist cannot do research without knowing what is going on in her field. This implies knowing which are the controversial articles, which are the ones he should take for granted, which are the ones he needs to read because of their relevance for his work and which are the ones he can skip because of their irrelevance or poor quality. A scientist needs to rely on representations about these cognitive productions; he needs to assess their cognitive worth, and in order to do so, he uses reputation. In practice, researchers — as well as Web users — will devote their cognitive resources to the sources of information that are readily available. These may be the scientific journals in their library, the articles that appear to be pre-eminent because often quoted, the authoritative scientists that are invited as 'guest speakers' at numerous conferences, etc. One can thus see that the scientific practice of search for information from others is not much different from the practices on the Web.

What is the rationale for devoting cognitive resources to readily available information? The cognitive foundation of this behaviour is straightforward: scientists, as lay people, work under severe cognitive constraints and limits, they just cannot read and understand the entire scientific production, but need to limit themselves to what is worthwhile for them. How do they determine what is worthwhile information for them? We are again facing the problem of information search decision making, and the solution is the same: cognitive attention is given to salient and available information. The question is also the same: What makes this behaviour rational, adapted to the exigencies of scientific thinking? And the answer is the same: the institutional setting makes it rational. And this institutional setting, again, is formed by a distributed assessment system.

The production of evaluative representations in science involves a distribution of cognitive tasks among numerous cognitive agents and processes. For instance, the evaluation of a job applicant's credentials is a representation that is determined by several factors including teachers' past evaluations of the applicant, his past employers and journal's referees (since publications constitute credentials). From one's education to the peak of one's career, expertise is constantly assessed. Grant applications, submissions of papers and job applications are part of a scientists' everyday life. The evaluation of a scientist could not be brought about by a single person: every scientist goes through a very large number of assessments, which stretch over a whole career, and may require different kinds of very specific expertises. Thus, the evaluative process in science is fundamentally distributed. The distributed aspect of scientific evaluation is even more apparent in the assessment of research results. The authority of an article crucially depends on its reception by the scientific community. It is through the judgements of other scientists that representations of the quality, relevance and importance of articles emerge.

The processes through which scientists and institutions assess research rely on a positive reputation system, i.e., a set of procedures that attach to actors or objects an attribute "that signals that they are more likely to be desirable for some sort of interaction than those without the attribute" (Whitmeyer 2000). In scientific practice, the positive attributes include holding a Ph.D. and research positions in more or less prestigious institutions (for people), being published in well known scientific journals and being widely quoted (for articles), or being ranked in the top first universities or most active research labs (for institutions). The pervasiveness of the reliance on indicators of quality is made apparent in Boix Mansilla and Gardner's sociological research on the assessment of scientific research:

> Faced with the task of making their assessment criteria explicit, researchers typically referred to indirect or field-based measures of quality. They pointed to indicators such as the number of accepted patents, publications, devices, and citations stemming from the work; the prestige of the universities, funding agencies, and journals in which it is placed; and the approval of peers and a broader community. "Simply counting things are easy answers as far as

> I'm concerned," claimed Jonathan Rosen, Director of the Office of Technology Implementation at CIMIT. "How many patents have you filed? How many patents have been licensed? How many new companies have been started? How many Science papers? How many Nature papers?" Field-based measures of this kind [...] [rely] on social procedures of peer review, inter-subjective agreement, and ultimately consensus as generators of acceptable insight. Our subjects [...] described these criteria as the standard way — however flawed — in which the quality of interdisciplinary work is determined at the forefront of knowledge production today (Boix Mansilla and Gardner 2003).

The positive reputation systems of science allow scientists to know whom and what to defer to, whom to trust *without further investigation*: it allows single elements to *reduce the cost* of trustworthiness assessment orienting their decisions. The positive reputation system of science tags sources of information as worthwhile: these sources of information are more likely to be desirable for cognitive interactions leading to fruitful scientific research.

The similarities between Web and scientific distributed assessment system are not so surprising. In fact, the reputation system implemented by search engines is inspired by citation analyses of scientific articles — an assessment procedure developed by Garfields in the 50's. Links in the Web and citations in the scientific literature are interpreted similarly as expressions of cognitive worth. Yet, the existence of distributed assessment systems is not restricted to these two cases. Distributed assessment systems are more widespread and comprehensive: in scientific practice they also include the continuous assessment of life span careers in addition to cite number of publications. When one goes to a medical doctor, one also relies on some distributed assessment system that issued the label of 'medical doctor'. Information about material goods is likewise produced by distributed assessment systems, at least insofar as consumers contribute to the reputations of brands. And the reputations of brands, in turn, determine consumers' choices. On the Web, commercial Web sites such as e-bay and Amazon also implement, with great success, distributed reputation systems.

4.2 General properties of distributed assessment systems

The existence of distributed assessment systems for information search can be explained functionally: they answer to an essential cognitive need — choosing to which source of information to devote cognitive effort. Distributed assessment systems perform a cognitive task — producing adequate representations of worthiness — that could not fruitfully be pursued by isolated individuals.

It would be interesting to investigate further into the generation of the institutional frameworks that produce distributed assessment system. Let me just repeat that the core institutions of such distributed assessment systems are reputation systems: they process information from multiple sources and produce a representation of cognitive worth. Such institutions can be found in Academic institutions and other reward systems. One may even see 'institutions' such as 'gossiping at the end of Sundays' religious rituals' as implementing reputation systems: gathering and redistributing (after some processing!) information about the reliability and trustworthiness of others. Of course, the functional aspect of these institutions is not by itself the cause of their existence, but it should certainly figure in the economic and social histories of their emergence.

So what is new with the Web's distributed assessment systems? Neither the function these systems implement nor the principles of the processes implemented are new. The change is to be found in the way Web reputation systems operate: they are *automatised*. It is difficult to pin down what exactly this automatisation brings about. It certainly brings a cognitive power that was never attained before. Few institutions, indeed, are able to centralise and process as much information as search engines' algorithms. This cognitive power is also manifest in the responsiveness of the system. It is continually and rapidly updated. A rapidity that proves essential for grasping the dynamics of the Web. So a change in cognition induced by technologies of Web search is certainly a quantitative jump in cognitive power. Qualitative change may come out of the quantitative growth. At the moment, however, I find it difficult to have a dichotomy of pre-Web and Web eras in terms of with vs. without filters, or dynamic vs. expert based assessment — as in the Eco–Casati debate.

The characterisation of the role and place of search engine technologies in the Web's distributed assessment systems allows pinning down some of its essential characteristics as common to any distributed assessment system for information search. These are:

- The distributed and cognitive aspects of these systems: the systems deal with information, and the processes are distributed among multiple agents, who issue multiple value judgements. The systems qualify as distributed cognitive systems. As assessment systems, they rely also on inferential assimilation of representations on the value of pieces of information and the worthiness of their sources.
- In order to solve the problem linked to the cognitive management of cognitive resources attributed to others as sources of information, the output of the system needs to combine saliency and availability within their

representational format. The representational medium use a 'less is more' effect in cognition (Goldstein and Gigerenzer 1999), which means that individual cognition benefits from having less information (a short list of URLs more reliably directs cognitive attention than a large set of URLs coupled with detailed accounts of their worthiness). The distributed assessment system structures the environment in such a way that there is a *fit* between the environment and the cognitive processes, which automatically directs attention and effort towards salient phenomena (Sperber 2005). The fit, in our case, is between naturally processed because available, and cognitively worthy — it leads to the maximisation of cognitive resources in one's search of information from others.

- A central problem of distributed assessment systems is to maintain this fit: to make sure that the most available sources of information are actually the most worthwhile. This is the condition for the maximisation that is associated with meritocracy: the ones who are listened to are the ones that deserve the most to be listened to. Thus, the worries about distortion and unintended effects should be central issues for reputation systems. One can promptly note the incentive to 'cheat' reputation systems, i.e., to bypass the goals of the system so as to make use of its actual functioning. But there is an opposite interest for the systems themselves that need to keep the cheats away so as to maintain their reliability and efficiency (and thus their existence). Maybe an unintended effect of search engines is to further the preferential attachment behaviour that structures the Web. Preferential attachment is the fact that links are mostly directed towards those URLs that have already been linked to. Since search engines direct attention towards those latter sites, they increase the probability of further linking to them. The unintended consequence is an over hegemonic structure and a lowering of the chances of newcomers to get 'chosen' and linked to.

A last property of distributed assessment systems is relevant to the question of change brought by technologies in the distributed cognition of communities. The representations' output of these systems, reputation labels, can be, and actually are, processed in combination by the end users. This means that I can use the results of a search engine for directing my attention and efforts, but I will also generally double check the results with some further information on the sources of information. For instance, I may use my knowledge of the prestige of authors when choosing among the ten first results of a SERP. Furthermore, this prestige is the representational output of some other distributed assessment system, such as the after-a-conference-day-gossiping or the prestige of their institutional affiliation (which may itself be the results of some (distributed) evaluation of public educational institutions). Because all these representations of worth are finally combined in the head of end users, they eventually influence each other. The hierarchy established by one reputation system feeds in the other systems' hierarchies. The assessments that individuals produce ultimately operate across reputation systems: when the assessment is taken into account by a system, then it shall (partially) determine the behaviour of some users that will produce further assessments taken into account by other systems. Distributed assessment systems operate across systems.

With the advent of new publishing technologies and media, with the creation of technologically supported reputation systems, the cross-operability of distributed assessment systems is fully effective. At first, it is the old hierarchies supported by already existing reputation system that inform the yet-to-be-built hierarchies. This supports the thesis that there is much continuity in cognitive decisions managing the allocation of cognitive resources. And indeed, a simple look at the current history of prestigious sources of information shows that no real revolution has taken place.

5. Conclusion

I have emphasized the continuity between pre- and post- search engines for three different reasons:

Firstly, I wanted to balance discourses which assumes that the Web introduces dramatic changes. Secondly, analytical abstraction reveals the hidden structure common to all distributed assessment systems and thus provides a template through which actual changes come out. The existence of important local changes cannot be denied. I have noticed, for instance, that marketing strategies have to face very new challenges introduced by the new stake of 'having one's web site visited'. With search engines, reputation may also free itself, to a certain extent, of the biases introduced by the interests of those institutions that implement their own reputation systems favouring certain criteria rather than others. This is because the Web combines information coming from different kinds of sources, that would not have communicated before. Also, search engines implement a radically new way of carving information: while information is traditionally structured in hierarchical typologies (e.g., discipline, sub-discipline, and field), search engines structure it with key-words and sets of key-words.

The third reason why I have emphasized continuity has to do with the relation between technological innovation and cognition. Often (as in a Marxist perspective) technological innovation is shown to be the source of changes, social or other. But the converse determination, the social and cognitive determination of technical innovation, should not be ignored. Analyses of the complex network of historical causes and determinations between technology and cognition have focused on different causal chains, as for instance:

Technology → practices (using cognitive artefacts) → (situated) cognition
Technology → modes of production → social structure → (distributed) cognition

But any technological innovation takes place within society at a given time and place. Technological innovation is thus itself situated in a given culture and social structure. Before it can act upon culture and society, an innovation arises from the current beliefs and goals of a community. In order to be successful technological innovations must 'fit in' the social and cultural situation. What 'fit in' really means is maybe the most fundamental question of the history, sociology and philosophy of technology. In this essay, I have introduced the cognitive dimension of this 'fitting in', viz. the need to manage cognitive effort in information search. Moreover, I suggested that the existence of distributed assessment systems is pervasive across cultures — i.e., distributed assessment systems would be a universal cultural answer to a universal cognitive problem. Of course, this does not mean that there is a single cultural solution to the problem of information search. Distributed assessment system is just an abstract notion, which is implemented in many different ways. Web users + search engines would be just another avatar of this cultural universal. But the notion of distributed assessment systems sends us back to some basic facts of social life: there are incompetent people and there are liars. But rather than simply avoiding collaborating with and trusting others, humans have found another solution, which paradoxically implies even more trust and collaboration: humans interact and collaborate in order to solve collectively the problem of information search. Here, the focus is on the causal relations that go from human nature, especially in what concern human's cognitive apparatus, to culture, which includes technical innovations.[6]

Notes

* I thank Roberto Casati, Gloria Origgi, Dario Taraborelli and Dan Sperber for their help and encouragement in the general project of developing an analysis of the epistemic impact of the Web. Gloria Origgi and Dan Sperber also commented on an early version of this paper.

1. Technological advances make it easy to create documents: there are user-friendly editors of html, and the knowledge of the basic skills for web-publishing is spreading; acquiring a domain on the Web is much facilitated and the prices very affordable, when it is not free. Of growing importance are the means of publishing without going through the html and transfer protocols: blogs and the Wikis are such means, which are more and more widely used (for instance, it is technically very easy to write one's article on Wikipedia).

2. Questions such as: shall we promote the development of the Web? How and for which purposes? What means should be used to manage information on the Web? Shall we implement censorship? — have led to a divide between Internet enthusiasts and Internet skeptics.

3. From Google's web site: http://www.google.com/technology/, accessed on 1 July 2005. For an account of the PageRank algorithm and its underlying mathematics, see Brin and Page (1998).

4. It is difficult to know what the algorithms of search engines exactly do, since they are kept secret for commercial reasons. Supposedly, the Hypertext-Matching Analysis factors in the texts that link to the document in question. It may also use some cluster analysis: the analysis of closely hyperlinked sets of documents.

5. Truth and aesthetic claims contain claims of universal validity.

6. The causal relation that goes from the nature of mind to culture has been emphasized by, e.g., Sperber (1996) and Tooby and Cosmides (1992).

References

Boix Mansilla, V. and Gardner, H. 2003. "Assessing interdisciplinary work at the frontier. An empirical exploration of 'symptoms of quality'". In *Rethinking Interdisciplinarity*, http://www.interdisciplines.org/interdisciplinarity.

Brin, S. and Page, L. 1998. "The anatomy of a search engine". http://www-db.stanford.edu/~backrub/google.html.

Casati, R. 2003. "Ce que l'Internet nous a appris sur la vraie nature du livre". In G. Origgi and N. Arikha (eds), *Text-e: le texte à l'heure de l'Internet*. Paris: Editions de la Bpi Paris, 51–75. [http://www.text-e.org]

Clark, A. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford : Oxford University Press.

Eco, E., 2003. "Auteurs et autorité". In G. Origgi and N. Arikha (eds) *Text-e : le texte à l'heure de l'Internet*. Paris : Editions de la Bpi Paris, 215–228 .

Goldstein, D.G. and Gigerenzer, G. 1999. "The recognition heuristics: How ignorance makes us smart". In G. Gigerenzer, P. Todd, and the ABC Research Group, *Simple Heuristics that Make Us Smart*. New York: Oxford University Press, 37–59.

Hutchins, E. 1995. *Cognition in the Wild*. Cambridge, MA: The MIT Press.

Sperber, D. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.

Sperber, D. 2005. "Modularity and relevance: How can a massively modular mind be flexible and context sensitive". In P. Carruthers, S. Laurence, and S. Stich (eds), *The Innate Mind: Structure and Content*. New York: Oxford University Press.

Sperber, D. and Wilson, D. 1995. *Relevance: Communication and Cognition*, 2nd ed. Oxford: Blackwell.

Tooby, J. and Cosmides, L. 1992. "The psychological foundations of culture". In J. Barkow, L. Cosmides, and J. Tooby (eds), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press, 19–137..

Whitmeyer, J.M. 2000. "Effects of positive reputation systems". *Social Science Research* 29: 188–207.

Author's address

Christophe Heintz
1 bis av. Lowendal
F-75 007 Paris
France

christophe.heintz@gmail.fr
http://christophe.heintz.free.fr

About the author

Christophe Heintz is working at the Institut Jean Nicod on the relations between cognitive and social factors of knowledge production. His research is focused on methodological questions in science studies, on the study of mathematical and anthropological knowledge, and on the role of the Internet in knowledge production. He has been trained in mathematics and philosophy at the Universities of Paris and Cambridge, and is currently completing his Ph.D. at the Ecole de Hautes Etudes en Sciences Sociales (Paris).

Speech transformation solutions*

Dimitri Kanevsky, Sara Basson, Alexander Faisman,
Leonid Rachevsky, Alex Zlatsin, and Sarah Conrod
IBM Thomas J. Watson Research Center / Cape Breton University
(Conrod)

This paper outlines the background development of "intelligent" technologies such as speech recognition. Despite significant progress in the development of these technologies, they still fall short in many areas, and rapid advances in areas such as dictation are actually stalled. In this paper we have proposed semi-automatic solutions — smart integration of human and intelligent efforts. One such technique involves improvement to the speech recognition editing interface, thereby reducing the perception of errors to the viewer. Other techniques that are described in the paper are batch enrollment, which allows the user to reduce the amount of time required for enrollment, and content spotting, which can be used for applications that have repeated content flow, such as movies or museum tours. The paper also suggests a general concept of distributive training of speech recognition systems that is based on data collection across a network.

Keywords: semi-automatic solutions; automatic speech recognition; batch enrolment; distributive training; editing interface

1. Introduction

As bandwidth for web and cell-phone applications has improved, there has been a corresponding increase in audio and video data available over the web. This proliferation of data has created new challenges in accessibility such as the need for better methods of transforming spoken audio into text. The creation of textual transcripts ensures that multimedia materials are accessible not only to deaf and hard of hearing users, but also to users who are "situationally disabled" with limited technical resources. For example, users with low bandwidth transmission can read text streams when full bandwidth video is not an

Pragmatics & Cognition 14:2 (2006), 411–442.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

option. Written audio transcripts are also prerequisites for a number of other important processes, such as translating, summarizing, and searching.

Human transcription of speech-to-text is expensive and requires trained experts, such as stenographers. Through the use of a human intervention called "shadowing", "parroting", or "re-speaking", automatic speech recognition has been used for live transcription. In these methods, skilled speakers train speech recognition software to recognize their own voices. These speakers then "shadow" the speech of untrained speakers, like in the process used in simultaneous translation. Studies suggest that the accuracy levels of software trained on such skilled voices make "shadowing" a viable tool for live subtitling (Lambourne 2004). Although effective, shadowing is a task that requires skill and training; therefore it is not widely available or affordable.

Automatic speech transcription is steadily improving, with error-rates dropping by as much as 25% per annum on specific data domains (Nahamoo 2005; Saon 2005; Pallett 2003). Nonetheless, full transcription of audio materials from all domains remains a distant goal. For example, current automatic transcription accuracy for broadcast news is approximately 80% (Saraclar 2002; Kim 2005). There is an obvious gap between current automatic performance and the professional requirements for captioning. As a result, expensive human intervention, rather than speech technology, is currently used in commercial captioning. This costly process poses another problem: Most audio and video information remains untranscribed, untranslated, unsummarized, and therefore unable to be searched.

Speech captioning is an important issue in our information society. This paper suggests ways to provide speech captioning by combining what speech automation technology can currently handle with what can best be handled through human mediation. The challenge before us is to utilize the human component most efficiently and most cost-effectively, while simultaneously improving speech automation technologies. This paper focuses on distributed post-editing as a viable mechanism to counter shortcomings in speech technology. The paper also describes several other methods to address current shortcomings of speech recognition, such as content spotting, batch enrollment and automatic speaker identification.

2. Large vocabulary speech recognition: Inflated expectations

The applications and anticipated success of automatic speech recognition (ASR) has had appeal for many businesses and industry sectors. Predictions over the last 20 years have positioned speech technology as a "game changer" for the business world, resulting in revenue opportunities in the billions of dollars. Those following the speech technology market can trace a long history of great expectations, which have not been matched by revenue or cost savings opportunities. For the general public, the first major foray into speech technology were dictation systems; and the results did not meet user expectations. This section traces the evolution of speech dictation in the marketplace, tracks some of the key missteps, and describes the ramifications of this experience for other speech technology applications.

One of the circumstances that enabled the creation of dictation recognition systems was the increase in the speed and memory of personal computers that had been taking place over a number of years. By 1995, PC systems on the market offered anywhere between 16 megabytes and 64 megabytes memory, and by 1995, systems had speeds of 100 and 200 MIPS.

The technology enabling dictation existed by 1990. Performance of the dictation recognition systems was significantly improved by having a speaker enroll or train the system. By reading a relatively small number of sentences, requiring 30–40 minutes, users could significantly improve the speech recognition system's performance. Dramatic improvements in speech recognition are also achieved by taking the output of the speech recognizer and feeding it back to the acoustic model. This method is called unsupervised adaptation (e.g., Woodland 1996). A third way to improve the performance of the speech recognition system is by updating the language model. In speech recognition, a language model refers to a probabilistic distribution capturing the statistics of the generation of a language, and attempts to predict the next word in a speech sequence. Adjusting language model probabilities in a language model is usually done by processing specific textual corpora that represent a speaker's topic or domain. This language model adaptation reduces a potential mismatch between the language model that a speaker started with and the language that the user is actually dictating to the system.

There was also an increase in options that users had in the dictation interface that facilitated on-line error correction. The user interface error corrective technique included such features as voice command control for navigating through a decoded text and displaying a list of decoding alternatives for words in the decoded text. A user could point to a word displayed in a dictation window by moving the cursor across the screen via voice commands (for example, commanding: "go back two words") and then choosing a correct word in a list of displayed alternatives. Another error correction feature allowed the user to repeat a word that was not decoded correctly several times. This method proved useful in helping to correct incorrectly decoded words. Experienced speech recognition users learned to choose appropriate error correction techniques for particular correction tasks, and significantly reduced correction time in comparison with novice users (Karat 2000).

Some speech recognition systems improved the user's acoustic and language models incrementally when the user corrected speech recognition errors that occurred during dictation. With such adaptation systems, the number of decoding errors was reduced during dictation sessions. Though errors still existed, they usually stabilized at word error rate levels ranging from 15% to 2–3% for different users who had in-depth experience with dictation systems (Bain 2005).

Adaptation capability, combined with the newly available speed and memory to run dictation on most PC systems, enabled the creation of the first set of dictation products in the mid 1990s. By 2001, over two million copies of Via Voice were preloaded and shipped on new IBM PCs. Despite the wide availability of speech dictation systems and relatively good decoding accuracy (many users could get above 90% decoding accuracy after training), the use of speech recognition systems did not become pervasive and in fact declined.

The providers of the time also exaggerated the performance of these systems. Claims were made about systems achieving 99% accuracy and handling two million word vocabularies. The 99% target was, as it turns out, rare and difficult to achieve. A system with a two million word vocabulary is not useful if, in fact, it is still getting a large portion of the words wrong. The overstated claims associated with dictation recognition had a deleterious effect on the product since users expected high recognition results and were disappointed that they could not in fact achieve this level of accuracy. There were also high expectations that the dictation system could help older users who were not skilled typists. But older users had difficulty learning to use the relatively complicated interface in speech recognition systems. Another factor affecting the usability of dictation systems was that users "type" thoughts differently than they "speak" these thoughts. Therefore, even if the user's thoughts, expressed via dictation, had perfect accuracy of the spoken word, users were still disappointed by the quality of the textual documents produced by the speech recognition systems.

All of these factors put a damper on the use of speech recognition systems for dictation. The speech recognition community began to focus instead on developing telephone applications and embedded speech recognition for command and control applications. Communities that depended on high bandwidth, large vocabulary speech recognition systems, however, lost access to a critical function. Large vocabulary speech recognition systems can provide textual access to speech for deaf or hard of hearing users. Additionally, dictation systems provide input options for motor impaired individuals and the needs of these communities are unmet with systems created for the telephony or the embedded market.

3. Using existing speech recognition technology for accessibility

3.1 Liberated Learning

The Liberated Learning (LL) consortium (Bain 2005; http://citizensforltc.org/issues_boomers.htm) is one example of how people use automatic speech recognition to caption lectures.

LL was launched in 1998 by Saint Mary's University (SMU) in Nova Scotia, Canada. SMU proposed a project to create a more fully accessible learning environment. A team at the university's Atlantic Centre of Research, Access, and Support for Students with Disabilities envisioned a paradigm that would "liberate" students from traditional, intermediary supports. LL specifically addressed a problem of providing access to lecture information. The research carried out to date within this program reveals important implications for students with disabilities and other stakeholders. Pedagogical results reported by instructors are equally intriguing. More than a dozen universities worldwide (in the US, Canada, China, Japan, and Australia) have joined the LL consortium.

In LL courses, instructors use a specialized speech recognition application called IBM ViaScribe. When used during class, ViaScribe automatically transcribes the instructor's speech and displays it as text for the entire class to read. After class, students can access ASR-generated course notes on the internet, available in various accessible formats: searchable transcripts, synchronous multimedia, and digital audio. The challenges using "standard" speech dictation software were obvious: the words needed to be accurate and readable, and students needed to be able to access the notes. Initially, the digitized text contained no sentence markers to distinguish independent thoughts. Text would appear word after word, line after line, in a continuous stream that quickly enveloped the screen. The requirement to verbalize markers such as "period" or "comma", as was required in desktop speech recognition systems to break up displayed text, was obviously not conducive to a lecture setting.

LL consortium participants have provided important usability feedback and requirements on ViaScribe user interface features and speech recognition components. As one of the outputs of the Liberated Learning project, a list of required functions to make speech recognition technologies suitable for real time transcription applications was identified (see below). In this list we indicate which of these features have been implemented, and which ones are currently being implemented and tested.

Feature	Comment
The real time speech recognizer should allow the lecturer to talk naturally, without interjecting punctuation marks.	This feature has been implemented by eliminating punctuation marks and employing user determined pauses to separate parts of texts.
When the speaker pauses, the text should skip to a new line (or some other separator like "…." should be printed), making the display more readable.	This feature has been implemented in ViaScribe. Separator features like a new line, period or other can be customized by users.
When the speech recognizer makes errors, it should offer viewers the opportunity to see the errors written out phonetically, instead of presenting an incorrect word.	This is implemented, for details see (Basson 2002, 2004).
The speech recognition system should limit the need for user training by working in speaker independent mode.	ViaScribe allows a user to bypass the training phase and use a speaker independent model. However, the decoding accuracy is usually improved after adaptation of a user model to a specific speaker.
The system should allow non-intrusive enrollment of speakers who do not want to spend hours reading sentences to explicitly create their speaker models.	This is implemented in the batch enrollment module in ViaScribe.
A user model should reflect a user's natural manner of speaking during their presentations, rather than creating user models best suited for dictation and "read speech" which is the result when users simply read enrollment stories.	The batch enrollment module in ViaScribe allows the capture of a user's natural speaking manner. It is being tested by LL partners. The result of the testing will be described elsewhere. Preliminary results indicate that training a user model from audio data captured during his/her presentations provides better decoding accuracy than a standard enrolment via dictation.
The system should automatically synchronize audio, captions, and visual information (slides or videos) to create accessible learning materials.	This is implemented in ViaScribe.
The system should allow networking models enabling every student to read a transcript on their client devices like laptops or PDAs. This networking model should also allow users to perform additional operations on their client display, such as annotating or editing a transcription, and customizing displays to fit their needs (such as choosing fonts, increasing font sizes, choosing colors).	A server model of networking is implemented in ViaScribe and is being used at Alexander Graham Bell National Historic Site in Nova Scotia (see below). A different approach that uses peer-to-peer wireless network is implemented by Mike Wald (see Wald 2005).
It should offer an easy-to-use error correction system, for post-hoc editing.	This is a major issue that is being addressed jointly with LL and CaptionMeNow transcription providers (described in (CaptionMeNow 2004)) The paper discusses some of ideas and ongoing efforts to solve this problem.

It should offer a lecturer an easy way to correct the most egregious errors on-line (during a lecture). Lecturers typically should not be expected to watch a decoded transcript while they are speaking. But a lecturer may recognize from student reaction (say, they are laughing as they look at the display) that there is something distracting about the transcript and look on a screen. If there is an error that should be fixed rather than left on the display — the lecturer should be able to quickly overwrite this error using available fast error correction options.	The real time editing is being implemented in the Protitle Live System that enables creation of broadcast subtitles in all major languages, using ViaScribe technology (Ghez 2005).
Real time editing options for lecture transcripts should be considered that allow one or several editors to correct a lecturer's transcript during a lecture (without interruption). There can be several scenarios on how corrections are displayed on a lecturer screen or in viewers' client displays. In one scenario, the lecturer's transcript goes to an editor buffer first and is made available for display only after it has been corrected. This variant may cause an unacceptable delay in displaying the transcribed speech. In another scenario, a transcript is displayed with errors, but when words are corrected, they change their appearance. For example, a word appearing in a different color than the rest of the text would bring the corrected words to the viewer's attention. Editors, who could also be students attending the lectures, need not correct all errors but only critical ones that affect the meaning of a spoken phrase. In principle, one can consider a networking system where every student in a lecture can edit a transcript during a lecture. Using a voting system, if there are conflicts in editing, a final edit can be displayed on a lecturer's screen or in student client displays. One can also consider another editing option: errors are fixed during a lecture, but corrections are not displayed on a screen until final lecture notes are delivered to students.	Some of the components of this system are being implemented at IBM Research.

3.2 CaptionMeNow

Another example of speech recognition for transcription is the CaptionMeNow project (CaptionMeNow 2004) that was initiated at IBM in 2004 and uses IBM ViaScribe technology. The project focuses on providing transcription of numerous audio webcasts.[1] Most webcasts are untranscribed and therefore, not accessible or searchable. The absence of transcription also makes it difficult to provide other text related processes such as translation, summarization, simplification, annotation. In most the cases transcription of webcasts can be done off-line rather than in real-time.

CaptionMeNow includes the following provisions to make transcriptions available for those who need them with minimal possible cost.

a. *On demand transcription*
CaptionMeNow suggests a model where, upon request by a viewer, a transcription of a webcast can be generated. In one scenario there can be a CaptionMeNow button associated with the webcast. The viewer of an audio webcast can click on this button, request the transcription and specify when s/he would like to get the transcription. The audio is then processed through the CaptionMeNow transcription framework to produce a transcription that is associated with the webcast. This transcription can be produced as a separate text file or aligned with audio/video or webcast presentation slides. This transcription process is needed only for the first transcription request since the transcription will now be available to all follow-on users. CaptionMeNow also includes execution of the billing process to provide payment for transcription services. Various business models could be considered to provide payment. One such option is a corporate model in which the corporation makes a financial commitment to ensure that corporate websites are accessible. In this case, an "on demand" transcription model of CaptionMeNow would allow corporations significant cost savings, since many webcasts are not accessed by users who would need additional provisions, such as transcription, to access them. The CaptionMeNow concept is being implemented at IBM jointly with Cape Breton University in Nova Scotia, Canada for educational presentations posted on an internal educational IBM website for internal web lectures and podcasts.[2]

The model for transcribing these educational webcasts is as follows: Several thousand hours of archived educational webcasts that were created before the start of CaptionMeNow in 2005 are transcribed only upon request of individual users; when a user clicks on the CaptionMeNow link on the educational website. All educational webcasts created after 2005, however, are required to be transcribed prior to posting. The transcripts can be provided either by the people who created the educational presentations, or transcription can be requested through CaptionMeNow services. Since these services began, it has been observed that there have been fewer than 10 requests to transcribe educational webcasts that are stored in the archives. There have been, on average, requests for 25–30 hours per month of transcription of educational webcasts in 2005 (these requests are done by the creators of the educational webcasts). Similar trends are noted for podcasts. It is expected that the number of requests will grow with better awareness about these services and streamlining the process of ordering podcasts and webcasts to be transcribed. Nevertheless, it appears that the on demand process of transcription provides significant savings for the company because only a small percentage of users actually request these services. This results in considerable savings, compared with captioning all audio material a priori with no awareness of user demand.

b. *Distributed transcription framework*
Off-line transcription of audio can be organized through subsequent execution of some of the processes that are shown in Figure 1. The following list explains these sub-processes.

Processing audio files through automatic speech recognition. Speaker independent speech recognition can be used for speakers for whom speaker models were not created. In general, speaker independent models lead to a relatively lower decoding accuracy and seem to require more post-decoding editing. If there are multiple speakers, some having available speaker models, then additional speaker identification processes are needed to identify who is speaking. If a speaker is identified, and a user model for this speaker is available, then this user model is activated for the current audio segment. Otherwise, a speaker independent user mode is activated. One can also use speaker dependent models for a cluster of speakers, for example, gender dependent speaker models that were trained for a female and male population separately, or speaker models trained for people with accents.
Shadowing (voicewriting) process. Accuracy of transcription using shadowing depends on the skill of the voicewriter, the number of new words in the speech, which can later be added to the system's dictionary, and the quality of the speech being transcribed. It was reported that under certain circumstances the decoding accuracy 95- 98% was achieved using ViaVoice for a live-subtitling environment (Lambourne 2004).
Manual transcription. Human transcription of speech-to-text still remains a valuable option if the audio quality of webcasts is poor. Human transcription is typically outsourced to locations with relatively low labor costs.
Editing. Editing of speech recognition errors is only efficient for some transcription tasks if the number of errors is relatively low (below 20%) (Bain 2005), otherwise, it is more efficient to transcribe audio manually. Off-line editing can be provided by specially trained people trained in methods that speed up editing.
Alignment of audio and text. Webcasts with video and audio should be aligned with transcribed text in order for a viewer to be able to associate a transcription with the related video. If a transcription text is obtained via an automatic speech recognition system, then the audio is already aligned

with a decoded text. Tools for editing the decoded text that is aligned with the audio should be organized in such a way that corrected text remains aligned with the audio. If the transcribed text was produced manually, automatic means for text-audio alignment of the text should be used or the alignment of the audio with the text should be done while a writer transcribes the text (by mapping small segments of the audio that is being played with the pieces of text that the writer produces).

While speech recognition cannot be used alone to reliably caption random audio, our goal is to determine whether a hybrid of speech recognition plus limited human intervention can nonetheless reduce the time and costs associated with captioning. In our transcription work, we found that one hour of audio could be transcribed in 5 hours using shadowing plus editing, dictating into Power Point using ViaVoice. The same audio using manual transcription performed by experienced writers required 6 hours of effort. It is known (e.g., Karat 2000; Koester 2004) that correcting speech recognition errors is the main obstacle toward making transcription via speech recognition an efficient process. This suggests that with improvements to the ViaScribe editing interface, experienced editors would be able to further reduce the overall transcription time. Experiments with transcribers in Nova Scotia identified a number of usability problems in the editing interface that need to be resolved in order for speech recognition to provide significant enhancement to cost or time requirements of transcription. Some of these problems are enumerated below.

– *Re-alignment of a recorded audio with an edited text*
A speech recognition system usually provides natural alignment of a recorded audio with a decoded text (Kanevsky 1997). This alignment allows audio, which corresponds to words, to be played simply by clicking on these displayed words. When a decoded text is edited, however, missed words are inserted. A re-mapping of the written words with the audio is therefore required.

– *Editing aligned audio-decoded text*
An editor can play a decoded text and edit parts of it when he/she observes inconsistencies between displayed words and played audio segments. Then the editor needs to indicate the exact point within a text where he/she would like to introduce and make these changes. In principle there are many ways to perform these actions, via voice commands, using a mouse, editing shortcuts, displaying a list of likely alternative words, etc. This is an area where significant progress in reduction of overall transcription time could be made with a smart editing strategy and interface. Efforts were made within ViaScribe to

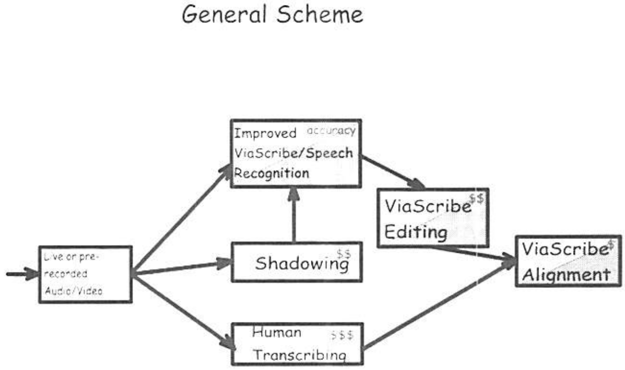

General Scheme
Improved accuracy ViaScribe/Speech Recognition
ViaScribe $$ Editing
Live or pre-recorded Audio/Video
Shadowing $$
ViaScribe $ Alignment
Human $$$ Transcribing

Figure 1. Paths for Captioning[3]

incorporate some of these suggestions for improving the editing interface, but since users suggest different alternatives such as different shortcuts for specific correction functions based on their own preferences, it is difficult to design the "universally desirable" editing interface that is optimal for all users. This problem is especially acute with professional manual typists who get used to some specific editing system and are slow adapting to the editing interface that is incorporated in ViaScribe. One of the options that is currently being investigated is to provide a broad range of editing tools to allow editors to change options dynamically to fit their individual needs (see Section 4.1).

c. *Grid implementation*

CaptionMeNow is being implemented at IBM using the Grid concept briefly described below:

– *What is Grid Computing?*

The Grid is a virtualized computer. Grid computing aims to create a single virtualized system by pooling undedicated, disparate resources. Typical resource classes include computational, network connectivity, and data storage. These resources are combined, potentially in parallel, to tasks such as the CaptionMeNow application and others such as processing batch jobs, and hosting web applications. The goal of an intelligent Grid is to (1) optimize the usage of these resources based on constraints such as locality, availability of free resources, and the business value, or "cost", of the individual jobs; and (2) create a shareable hosting environment for easier software installation and infrastructure maintenance.

- *Grid Architecture*

The Grid architecture is based on mixed UNIX environment of AIX and Linux operating systems and distributed shared network file systems. All applications reside in shared file systems. The Grid scheduler considers resource optimization such as CPU, memory and network bandwidth as local resource constraints, and the total number of instances for an application, resource requirements such as CPU, and memory as application constraints. We integrate IBM LoadLeveler in the Grid infrastructure.

- *Why use Grid for Caption-Me-Now?*

We use Grid Computing for the CaptionMeNow (CMN) application to get transcription results in a timely manner. The Grid allows users to run application tasks in parallel to increase the throughput. Users can submit on demand requests for transcription services using Grid resources.

When a Web page contains an audio file, the user can click the CMN button and send a request to the Grid to transcribe the audio file. The scheduler

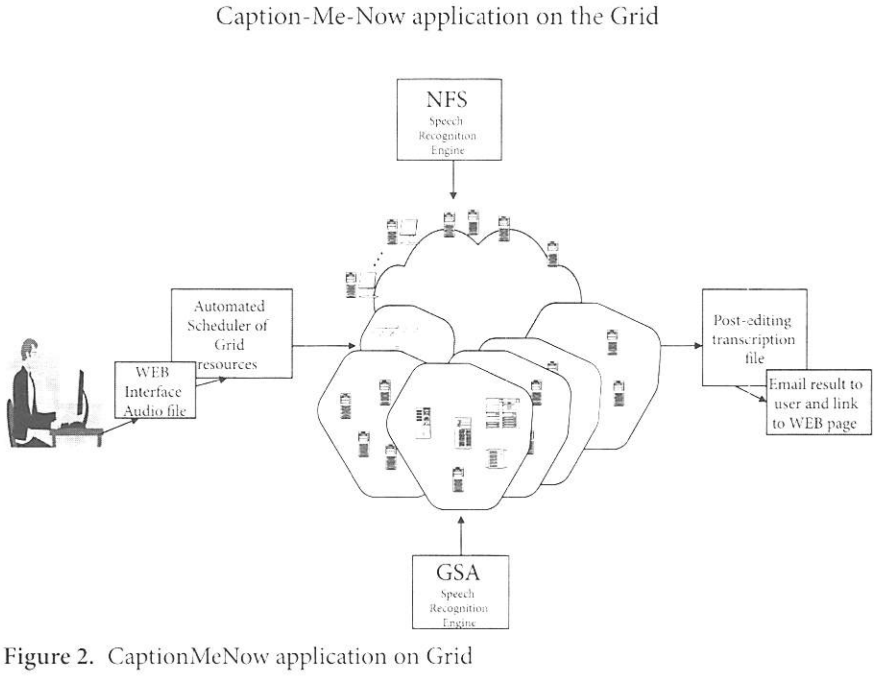

Caption-Me-Now application on the Grid
NFS
Speech
Recognition
Engine
Automated
Scheduler of
Grid
resources
WEB
Interface
Audio file
Post-editing
transcription
file
Email result to
user and link
to WEB page
GSA
Speech
Recognition
Engine

Figure 2. CaptionMeNow application on Grid

of Grid resources receives the request. We use the IBM Loadleveler product as the Grid scheduler. The scheduler generates a job to run on the available Grid computing nodes. The job invokes an IBM Speech Recognition Engine to do transcription. The Speech Recognition Engine resides in shared network file systems such as IBM Global Storage Architecture (GSA) or Network File System (NFS). The Speech Recognition Engine processes a job on one or many computing nodes and creates a transcription file. A service provider receives the transcription file for further post-editing. After the editing is complete, the service provider emails the file to the transcription database inside IBM and a link is generated on the web page indicating availability of transcription for the audio file.

3.3 Museum applications

Another successful application of ViaScribe speech recognition technologies was recently demonstrated at the Alexander Graham Bell National Historic Site (AGBNHS) in Nova Scotia (Conrod 2005). Conclusions emerging from this application are presented in the following section. The Baddeck Liberated Learning Showcase was launched in 2004 and was a project designed to explore how the use of real-time transcription could enhance interpretative talks within a National Historic Site. Due to their specific nature, museums provide an ideal environment to employ speech recognition technology for several reasons. One reason that speech recognition tools are particularly well suited to museum settings is because people that are hired to interpret museum stories tend to be highly proficient public speakers. Additionally, the content that they deliver is very focused, with a specific vocabulary of regularly recurring words. In general, tour guides repeatedly deliver the same content, modified for the group that is being addressed, day in and day out. Because of these factors their resultant voice profiles have proven to be quite strong, which has assisted with accuracy within the context of the interpretative talks that are being delivered.

Despite the targets achieved from the tour guides' strong voice profiles, museums present unique challenges that relate to displaying the transcribed audio as well as catering to the typical audience and environment in which material is presented. These issues required the development of techniques around the use of speech recognition and handheld devices. In public settings, small, unobtrusive delivery devices are crucial for high levels of user satisfaction and accessibility throughout a site. By using digital assistant devices such as cell phones and other handheld computers, visitors can move through a public setting such as a museum and refer to the transcription of information on demand, at their convenience and as frequently as they require. The availability of an on-demand accessibility device appears to increase user satisfaction as it allows for seamless accessibility while preserving the character of the museum and the anonymity of individuals using the adaptive technology.

Another issue is that museums cater to a diverse population. Due to the numbers of people that visit on a daily basis, there are myriad different learning styles, disabilities and special needs that can be effectively accommodated through the use of speech recognition software. Some of the issues that can be effectively met by blending speech recognition into traditional interpretative talks include accommodations for those with learning disabilities, persons with hearing loss, those with different learning styles, as well as techniques for people with native languages other than that of the primary speaker.

In addition to qualitative indicators for using speech recognition software in a museum, there are quantitative rationales that provide significant support for employing real-time transcription in interactive public environments. In Canada, reports indicate that one in ten people are classified as deaf or hard-of-hearing. In addition, last year twenty percent of the visitors to the Alexander Graham Bell National Historic Site were seniors. Older people often have hearing loss that can be successfully accommodated through the use of speech recognition technologies. As the population of baby-boomers entering retirement grows, the percentage of individuals requiring adaptive technologies is going to increase commensurately. According to the US Bureau of the Census, the number of baby-boomers aged 65 to 84 is expected to grow 80 percent in the next 30 years (http://citizensforltc.org/issues_boomers.htm).

Real-time transcription presents significant value for people whose native language is not that of the speaker. Museums often receive Japanese, Italian, Chinese, German and other visitors from around the globe. Research has indicated that although auditory comprehension may be low, non-native speakers often have high reading comprehension in English. Having text projection in conjunction with spoken tours can greatly increase the enjoyment and understanding for non-English speaking visitors. At any given time, it is almost impossible to accommodate this diversity of languages though real-time tours. By using speech recognition, techniques for real-time text display in a variety of languages in conjunction with the language that the tour guide is speaking can be employed. This unique process is made possible due to the nature of museum interpretation. Museums generally deliver repeating content on a daily basis; although the exact transcript may change, messages remain consistent. By employing special speech recognition tools, tour guides can use "content spotting" techniques to deliver text projection in a visitor's language of choice without knowing a word of the language that is being projected in synchrony with their interpretative talk. The system includes a pre-stored version of the tour guide's typical lecture, in a number of languages. When a tour guide presents the tour in one language, the system identifies the matching terms in the visitor's language ("content spotting") and displays the paragraph that corresponds to what the tour guide is saying at that moment, in the language of choice of that visitor.

Additionally, written language is sometimes easier for non-native speakers to understand. Research similarly indicates that there are many learning styles including auditory learners, textual learners, and visual learners that benefit from the availability of on-demand transcription. It has been found that some individuals, although not overtly needing real-time transcription, benefit from the opportunity to read information during museum tours and interactions.

This technology can also be contrasted with currently employed technologies such as pre-recorded audio tours which are common in many museums. These audio tours provide visitors with self-guided interpretative talks that highlight different objects on display as well as sections of the museum. Although these systems provide a comprehensive picture of the museum itself, they tend to create an isolated environment for the user, causing them to be cut off from others within the museum. As Bederson points out, this can cause visitors to feel disappointed as social interaction is one of the attractive aspects of the museum experience. Bederson (1995) states, "…part of the reason many people go to museums is to socialize, to be with friends and to discuss the exhibit as they experience it. Taped tour guides conflict with these goals because the tapes are linear, pre-planned, and go at their own pace". Additionally, Bederson point out that these pre-recorded tours address very specific sections of the museum and the delivery cannot be altered or changed based on a visitor's interests or questions.

The use of technology for real-time captioned tours can offer several advances to this traditional pre-recorded audio solution. One significant benefit is that the real-time captioning mirrors the words of the museum tour guide, allowing for a more dynamic experience. In addition, questions from the audience can be repeated and incorporated into the tour, allowing the group to interact and the guide to change directions to suit the mood of the group. A solution is also being implemented that will allow the display system to switch between content-spotting and speech recognition. For example, if the system detects that there is a bad match between speech recognition decoded output and a pre-stored version of the speaker's typical text, it will automatically switch to live speech recognition, adapting to the fact that the speaker apparently has gone "off script". Although the system will generally be displaying pre-stored text, it will still remain dynamic as the text being displayed will change to mirror the current talk that the tour guide is delivering.

Although there are many benefits to the use of speech recognition tools for visitors within a museum, it is important to analyze the impact on staff deploying the technology. Research has shown that one significant issue in the success of text-projection is the comfort level of the staff who demonstrate the technology. Museum guides typically have jobs that require no technical ability. When asked to demonstrate speech recognition technology, this traditional technology-free job becomes one where the tour guides are expected to utilize these tools while achieving a benchmark of 90 to 95 percent accuracy when animating stories to the same level that they would have without any technology responsibilities. It has been found that this expectation and their own desire to provide high-quality projections to visitors adds pressure to their jobs which, in some situations, has adversely impacted their accuracy levels in real-time animations. Because of these factors "content-spotting" is being used to assist guides in achieving near 100% levels of accuracy which removes much of their apprehension and provides increased confidence in the words that are being displayed to visitors.

Visitor feedback has also resulted in some interesting observations. Visitors with hearing loss have predictably given this technology very positive feedback. It is also of interest to note that many visitors who indicate that they do not have hearing loss, report that the use of real-time transcription during museum interactions enhances their visit and helps them follow stories and information better. This may be explained by the fact that as people age they often begin to experience low-grade hearing loss, which manifests itself by making isolated words difficult to hear or information difficult to process in large groups. Although they do not recognize that they have any degradation of their hearing, the technology is there and available to assist with small comprehension errors when they do occur. Additionally, the fully hearing textual learners seem to enjoy the novelty and availability of the text as a supplement to the interpretations. By having real-time transcription available they are able to use transcription as often as they like as it allows them to benefit from both auditory and visual input simultaneously.

The following section provides more detail on real time transcription for the Liberated Learning project and off-line transcription for CaptionMeNow.

4. Integration of human intelligence and computers

Any technology that requires artificial intelligence is imperfect. Therefore, we need intelligent integration of technology with human assistance. A single, well-equipped, trained and well paid person is more efficient than a large number of poorly trained and poorly paid coworkers. Often, quality cannot be achieved as the arithmetic sum of low quality operations. This is true for both humans and for computers. Despite the fact that software and hardware technology, developed by humans, is getting more complicated, computers are exposing their natural lack of intelligence. While a bare-handed human cannot compete with computers in the speed and accuracy of their calculations, the same human, trained and equipped with proper computer software, through the right interface, can outperform heavy weight computers as soon as the job requires slightly more intelligence. The 21st century will demand more and more of these well-trained intelligent and computer-powered individuals. The right software will be demanded and created for those individuals. The human–computer interface, versus current computer–human interfaces, will play a more and more important role as soon as the efficiency of this approach is effectively demonstrated.

Thus, a computer should be treated less like an independent thinking machine and more as an aid for the human who does the real thinking.

4.1 Editing innovations

Speech automation technologies can provide transcription, but editing work is typically necessary to ensure a high level of accuracy. In one case, as in interactive dictation, an author of the dictated speech corrects decoding errors. In another case where speech is uninterruptible, as in real-time captioning, editors correct speech recognition errors for other speakers. Editing will slow down the process, however, since multiple hours of editing might be required to perfect a one hour transcription. In some cases, real-time editing will be necessary to ensure that the correctly captioned material is immediately available. Multiple editors can be used to accelerate the editing process, but this presents another challenge: how to efficiently coordinate multiple editors, and ensure that the final product is provided in real time and appropriately synchronized. Off-line transcription of audio that is stored in webcasts or library archives requires a different approach for editing which depends on the accuracy requirements, which then depends on time and availability of speech recognition and human editing resources.

There are several issues that should be developed in order to increase the user interface editing efficiency.

- "Work Sharing" enhancements will enable multiple editors to be privy to what other editors have already done, and what sections other editors are currently working on.
- A special approach can be used to provide a "quasi-real-time" transcription, which should be produced within a small delay (like 5–7 seconds). Such delays of transmission of video are acceptable in TV broadcasts of live events. Video is delayed until transcription is produced and then transcription is inserted into a video stream. In this "quasi-real time" scenario, an expert (for example, a person who does shadowing) can mark speech recognition errors by clicking on them. The marked words are then split between editors who work on audio and textual segments that are assigned to them. This assignment of textual segments to different editors can be done by marking these segments with a different color.
- Editing tools will also include multiple input mechanisms such as a mouse, keyboard, touchscreen, voice, pedal movement, or a head-mounted tracking device capable of moving a cursor to a point at which the user is looking on the display. Additionally, different aspects of the task might be better handled by different interface tools. For example, "spotting" errors could be easier to handle with a head-mounted device. Repairing phrases might be best handled via a keyboard in the "quasi-real-time" scenario that was described above. Interactive dictation, for example, when the speaker edits errors himself, can be done mostly orally/visually, perhaps with just a pointer to highlight what is wrong and needs to be repeated. A quick mouse highlighter and oral feedback might be the fastest and most efficient model. The most efficient routing mechanism will be determined empirically based on task assessments. For off-line tasks such as transcription of audio archives, editing can be distributed randomly, based on availability of editors. Alternatively, the editing work can be distributed hierarchically. For example, the "first pass" review can be provided by "phase 1" editors, with a certain level of skill; the "second pass" review may be distributed to "phase 2" editors with a different level of skill. Some editors may have unique expertise in particular terminology; this phase of editing will be routed to these more specialized editors.
- The user interface of choice needs to be adaptable to the user, and changing dynamically depending on what types of errors are displayed. The different error types result in displays of different kinds of tools to facilitate editing

these errors. The tools should be multimodal, allowing the editor to choose the best modality for a given problem. A word that is identified as incorrect, for example, can then lead to a display of an alternative word list, so the editor can just click on the word that he wants to correct; or the reader can repeat orally, to try for a better recognition the second time. In other cases, it will allow the editor to use handwriting and to overwrite the incorrect word. The interface suggests user modalities that would fit the most suitable user for this type of error. The network of editors will also provide data that allows the system to draw conclusions about editor habits and preferences, thereby improving and customizing the editor interface across the network.[4]
- In current speech recognition dictation systems, there is no "look ahead" or "look behind" mechanism. A particular error may occur repeatedly, and each instance of error needs to be corrected individually. The system should instead automatically correct similar errors across the entire transcript. Basically, correction of an error should automatically modify language probabilities and thereby allow automatic correction of many other errors.

The distributed network of editors can benefit from distributive error learning. For example, the system can detect that an error was corrected by an editor in one place. The system can then learn from that situation and provide the correction automatically to another editor in another location. In this scenario, it updates speech and language models globally. In this sense, editing and training and learning and networking become a distributed task, with cognition shared across a network.

4.2 Usability enhancement: Batch enrollment

Standard speech recognition enrollment requires a speaker to read prepared stories aloud. The recorded audio is used to train the speaker model. A typical enrollment process involves reading one story and usually takes 20–30 minutes. Some users may read three or four different prepared stories to improve their acoustic models. Reading more than 3–4 stories rarely improves decoding accuracy. Further improvement in the user model may occur when the user corrects speech errors that occurred during dictation.

This enrollment process has several deficiencies. Many people are not willing to spend hours training speech recognition systems to achieve higher accuracy. Some are not willing to read even one enrollment story. Such situations can be observed in the medical arena. Medical doctors need to produce a lot of written material throughout the day. They usually dictate such materials to a tape that is then transcribed by medical transcription services. Such transcription services usually involve manual transcribers. A similar situation was observed in the Liberated Learning project, where only a handful of lecturers, "early adopters" of the speech recognition technology, agreed to enroll their voices into the speech recognition systems.

It is also difficult to enroll the system when there is no access to the particular speaker. Many official dialogues, speeches, and other audio files are not accessible because of limited access to train certain subjects. Another problem with the standard enrollment is that it records speech when a speaker reads a story rather than free conversational dialogue. An acoustic model that is created from a story read by a user may not capture certain characteristics of unconstrained speech.

This requires creation of non-intrusive enrolment, or so called "batch enrollment", that allows enrolling speakers indirectly without asking them to read enrollment stories. This concept is described in a recently issued patent (Kanevsky 2004), and there are several methodologies to incorporate this approach. In batch enrollment, the audio of speakers is recorded during their natural activities, for example, when they give speeches, participate in meetings, talk over the telephone, etc. This recorded audio is transcribed manually; for example, by using stenography or by using semi-automatic technologies such as a combination of speech recognition technologies and human assistance like shadowing and/or editors. Then the recorded audio and transcribed speech are aligned using automatic or manual alignment means and a speaker model is created for this aligned audio and transcribed text.

Batch enrollment is an important tool for gradually reducing the transcription cost of webcasts in the CaptionMeNow paradigm because much of the audio/video information available on the web is provided by the same speakers, repeatedly. For example, presidential addresses could be available through the Library of Congress. President Bush has not "trained" speech recognition systems, and yet there may be hundreds of hours of audio information in his own voice. If a user selects an address of President Bush for captioning, the first request may be routed through the "shadowing" path or the "stenography" path. The resulting audio-plus-transcription, however, can then be used to create a voice model for President Bush. As more of President Bush's addresses are selected, and more transcriptions are created, the voice model will incrementally improve. When the baseline speech recognition accuracy surpasses some pre-determined threshold, speech recognition can supplant shadowing or stenography as the preferred method to create captioned webcasts for President Bush, and further presidential addresses will be captioned much more cost-effectively.

The batch enrollment method as a way to efficiently create training models is currently being explored in the university lecture environment. The batch enrollment method has been evaluated in this setting as follows: Professors have digitized recordings of multiple lectures. The lectures are accurately transcribed, since they have been presented through ViaScribe and then edited as well by a teaching assistant. The existing transcribed lectures have then been used to create a training model for the professor, assuming that none had been created earlier. The "batch enrolled" accuracy models were then compared to the accuracy models based on the speaker independent (general) model. Results are presented in Figure 3. The lecture accuracy results on the batch enrolled lectures actually surpass the accuracy scores for speaker independent models and they were approximately the same as the accuracy results received when speakers performed the enrollment process. This result demonstrates batch enrollment as a useful method in creating acoustic models for speakers that did not go through the enrollment process. We expect that by adding more hours for the batch enrollment training and performing the enrollment multiple times, the accuracy decoding results will surpass the ones that were received via individual standard enrollments. In Figure 3, results are presented separately for a speaker that had achieved a high level of accuracy when a

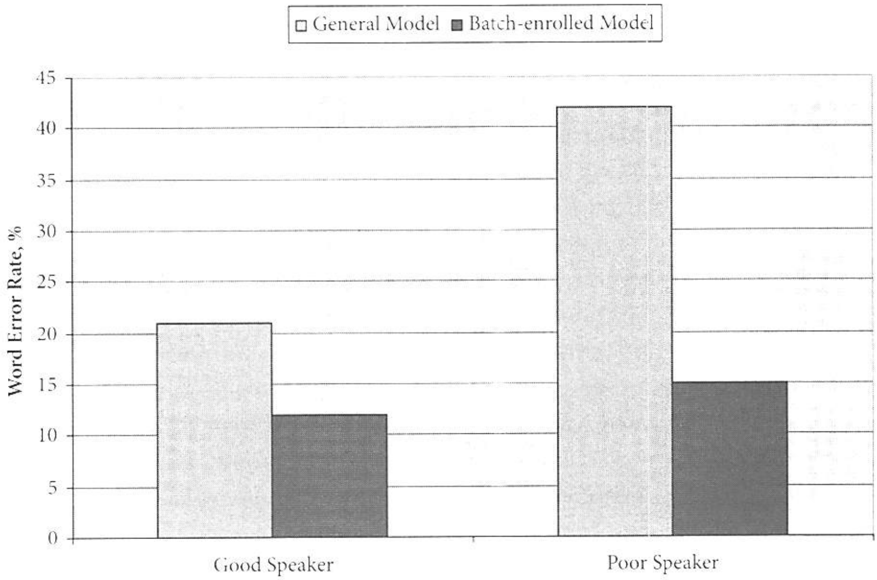


Figure 3. Impact of batch enrollment on ASR accuracy

general (speaker) independent model was used and a speaker that had bad accuracy in a speaker independent mode.

4.3 Training user models across a network

The amount of available training data is a critical factor for training user models that give high quality speech recognition results. Sometimes it is difficult for one person to generate as much training data as is needed to optimize recognition results even with the batch enrollment process.

Another solution to this problem of generating the sufficient amount of training data is to collect data from similar users that are connected over a network as was suggested by Kanevsky (2003). There are many methods for clustering speakers by similarity. One way to do this is to use the speaker identification technique that allows the identification of a speaker by voice patterns among a set of speakers. This technique uses various metrics to identify how close speech produced by speaker A is to voice pattern prototypes that characterize some speaker B. If there is a very close match, then speaker A can be put into one cluster with speaker B. If there is not enough acoustic data collected for speaker A to train a speech recognition system, and for speaker B a lot of acoustic data was collected, the acoustic data from speaker B can be used to train the acoustic user model for speaker A.

Adaptation methods exist that allow for the improvement in speech recognition by creating a user model that is trained on a combination of both types of data — the personal data collected from this speaker and data collected from other speakers that have some common characteristics with the original speaker. These common characteristics could include gender, age, language or accent. Clustering speakers according to their language or accent is most promising. In order to collect accent data across a network one needs to provide accent identification and to create an accent data base. A well known method of determining an accent from acoustic features is described in (Lincoln 1998). In this approach, a continuous speech recognizer is trained by several speakers with different accents. For accent identification an accent vector is extracted from each individual user's speech and the accent vector is classified. The accent vector is associated with each of the training speakers. The speakers are then clustered based on the distance between corresponding accent vectors, and the clusters are identified by the accent of the member speakers.

In general, the data from different users can be collected on a server and updated models can be sent back to local models. For each conversation that was recorded over a network, a speech waveform is captured and digitized, and acoustic features are extracted. The extracted features are passed to other connected systems and used to modify the speech recognition systems of similar users clustered in the same user cluster. The features are correlated with at least one attribute such as gender, age, accent, native language, dialect, socio-economic classification, educational level and emotional state. Some features, such as a speaker's dialect can be determined from the accent and the usage of keywords or idioms which are specific to a given dialect. For example, in the French language, the choice of *nonante* for the numeral 90 instead of *quatre vingt dix* would identify the speaker as being of Belgian or Swiss extraction, and not French or Canadian. Further, the consequent choice of *quatre vingt* instead of *octante* or *huitante* for the numeral 80 would identify the individual as Belgian and not Swiss. In American English, the choice of *grocery sack* rather than *grocery bag* might identify a person as being of Midwestern origin rather than Mid-Atlantic origin. Another example of Midwestern versus Mid-Atlantic American English would be the choice of *pop* for a soft drink in the Midwest and the choice of *soda* for the corresponding drink in the middle Atlantic region. In an international context, the use of *holiday* rather than *vacation* might identify someone as being of British rather than from American origin. The gender of the user can be determined by classifying the pitch of the user's voice, or by simply clustering the features. In the latter method, voice prints associated with a large set of speakers of a given gender are built and a speaker classification is then performed with the two sets of models. Age of the user can also be determined via classification of age groups, in a manner similar to gender. Although having limited reliability, broad classes of ages, such as children, teenagers, adults and senior citizens may be separated in this fashion. Native language of the user can be determined in a manner essentially equivalent to accent classification. Additional information about the native language of the speaker can be added to define each accent/native language model.

4.4 Content spotting

Many audio-based information streams have repetitive components, for example lectures, training, tourist information delivery. The repetitive aspect can be leveraged by preparing transcriptions or translations ahead of time. When a speaker says a particular "trigger phrase", the system can perform automated "content spotting" and thereby launch a prepared section of text, translated into the language of choice. This reduces the burden on the automation technologies (speech recognition, machine translation), thereby increasing the accuracy and efficiency of the proposed system.

Technology exists to synchronize spoken audio with a previously created script using speech recognition to align the two media. This technique is based on IBM issued patent (Kanevsky 1997). It suggests a process using a speech recognition system where an audio file search should be performed. Decoded text that the speech recognition system would produce is aligned to an audio file. The search for keywords could be done in a decoded text that is aligned with the audio. Spotting a key word in the text would allow identification of an audio segment associated with the key word. Clicking on the key word would play the audio segment. This search could fail to work if some key words in the text were decoded incorrectly. In order to overcome this problem one can search through phrases rather than words and match phrases that approximate the decoded text, by matching a few words in a phrase. For example, one can request a search by the phrase "CaptionMeNow". If there is a match between phrases in the decoded text by two essential words, then the decoded phrase can be produced as a search result. For example, a possible decoded phrase like "Caption *My* Now" would be found with these criteria. This concept can be extended to producing multilingual transcriptions of movies and live performances on PDAs as shown in the following usage scenario:

> *A French tourist in London wants to watch an English performance or movie. The French tourist goes to a website that stores translations of popular performances and movies, and downloads a time-stamped French translation of this performance to his PDA. During the performance, the PDA captures audio and sends it to a server. This server matches the audio and time stamped English transcript and sends a time onset for the French text in the tourist's PDA, to launch the onset of the translated text. The tourist can either track the synchronized translated text associated with the live performance, or listen via text-to-speech synthesis.*

This methodology can be extended to situations where the referenced content is not repeated exactly the same way from one iteration to the next, as is the case with lectures or training materials that are repeated, but not verbatim. A lecturer can repeat presentation material using slightly different words for the same content. Using topic spotting technology developed by research, we will spot equivalent content and time stamp it to trigger synthesized audio or pre-stored (multilingual) transcriptions corresponding to the same content spoken by the lecturer in previous deliveries.[5]

5. Summary

This paper outlines the background development of "intelligent" technologies such as speech recognition. Despite significant progress in the development of these technologies, they still fall short in many areas, and rapid advances in areas such as dictation have actually stalled. Reduced efforts toward developing speech recognition technology for dictation have inadvertently affected the development of speech recognition for high bandwidth applications. This negatively impacted the accessibility area, where high bandwidth, large vocabulary speech recognition technology was critical and required for applications like webcast captioning and lecture captioning. Speech recognition technology has launched new business areas such as speech applications over the telephone, and speech for embedded technology. This, however, did not fill the gap of speech recognition in high bandwidth arenas.

Alternative solutions like stenographic transcription address some of these gaps, and innovative efforts such as the Liberated Learning consortium advanced the state of the art for providing transcription in classrooms using speech recognition.

While the technology still needs to evolve, we have presented a very general concept for filling some of the gaps left by technology-driven solutions of speech recognition. We have proposed semi-automatic solutions — smart integration of human and automatic efforts. One such technique involves improvement to the speech recognition editing interface, thereby reducing the perception of errors to the viewer.

Other efforts underway make the technology easier to use for the speaker. Batch enrollment, for example, allows the user to reduce the amount of time required for enrollment.

Clever selection of application can also bypass some of the technical shortcomings. Applications that have repeated content flow, such as movies or museum tours, can more easily be transcribed, and even translated. For these areas, we have developed content spotting technology whereby key words or key phrases trigger the right content. Experiments at the Alexander Graham Bell Museum demonstrate the promise of this technology.

We also provided specific technical suggestions on how interactivity can be used to improve speech recognition errors in the middle of decoding and suggested a general concept of distributive training of speech recognition systems that is based on data collection across a network.

Important issues remain unresolved and should be the topic for future research. For example, there is a need to develop adaptive user interfaces for error correction that dynamically present the best tools for rapid correction that also make errors obvious. This includes development of an optimized scheme for error correction. For example, identification of errors in one part of the text can automatically trigger identification or correction of similar errors elsewhere in the transcript. Another important area that should be developed is the infrastructure for distributing process components across a range of services to the appropriate technologies and human providers, and finding the best path for efficient, cost-effective allocation of work.

Notes

* We would like to thank Allen McCormick, Keith Bain, Stanley Chen and other reviewers of this paper for their insightful suggestions and assistance.

1. Webcasts are defined by Merriam-Webster's Online as "a transmission of sound and images (as of an event) via the World Wide Web" that are currently available online. Merriam-Webster Online definition of webcast http://www.m-w.com/dictionary/webcast

2. TechEncyclopedia defines a podcasts as, "(a) recording of a non-music audio broadcast (news, sports, discussion, etc.) in the MP3 format for playback in a digital music player". TechEncyclopedia Definition for podcasting http://www.techweb.com/encyclopedia/defineterm.jhtml?term=podcasting

3. Dollar signs in blocks in Figure 1 show relative "costs" of a corresponding process. For example, human transcription with a stenographer is generally more "expensive" than shadowing. These situations were observed for some off-line CaptionMeNow transcription processes that were tested jointly by IBM and partners of LL in Nova Scotia. The diagram in Figure 1 does not cover all possible hypothetical transcription processes that would require a more elaborate analysis. For example, for some kinds of audio data, manual transcription that is performed in India would be significantly cheaper than similar transcription work performed in western countries. In our preliminary experiments, outsourcing of transcription of webcasts to India did not provide a viable long term solution because of reliability issues such as long turnaround times, higher error rates for transcription, and communication issues such as difficulties in transferring large volumes of audio data.

4. In Appendix 1 we describe statistical methods that could be used to create an adaptable editing UI.

5. Appendix 2 presents a brief description of the algorithm for content spotting that was implemented at the Alexander Graham Bell Museum for multilingual transcription of museum tours.

References

Bain, K., Basson, S., Faisman, A., and Kanevsky, D. 2005. "Accessibility, transcription, and access everywhere". *IBM Systems Journal* 44(3): 589–603.

Basson, S., Faisman, A., Ferre, W., Ghez, J., Kanevsky, D., and Quinery, J. 2002. "Lipcom: Speech recognition as a teaching aid for hearing-impaired children". In *Proceedings of the Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment* (CVHI'2002), Granada, PAGES.

Basson, S., Kanevsky, D., and Maison, B. 2004. "Hierarchical transcription and display of input speech". US Patent 6,785,650.

Bederson, B. 1995. "Audio augmented reality: A prototype automated tour guide". *Proceedings of CHI '95*: 210–211 [Bell Communications Research].

Berger, A., Della Pietra, S., and Della Pietra, V. 1996. "A maximum entropy approach to natural language processing".*Computational Linguistics* 22(1): 39–71.

Conrod, S. 2005. "Enhancing accessibility in interpretative talks". In *Proceedings of the Conference on Speech Technologies: Captioning, Transcription and Beyond.* IBM. http://www.nynj.avios.org/Proceedings.htm

Ellozy, H.A., Kanevsky, D, Kim, M.Y, Nahamoo, D., Picheny, M.A., and Zadrozny, W.W. 1997. "Automatic indexing and aligning of audio and text using speech recognition". US Patent 5,649,060.

Ghez, J. and Gue, J. 2005. "Protitle live — ViaScribe for broadcast subtitling". In *Proceedings of the Conference on Speech Technologies: Captioning, Transcription and Beyond.* IBM. http://www.nynj.avios.org/Proceedings.htm

Kanevsky, D., Zlatsin, A., and Pickover, C. 2003. "Educational monitoring method and system for improving interactive skills based on participants on the network". US Patent 6,505,208.

Kanevsky, D., Basson, S., and Fairweather, P. 2004. "Integration of speech recognition and stenographic services for improved ASR training". U.S. Patent 6,832,189.

Karat, J., Horn, D., Halverson, C, and Karat, C. 2000. "Overcoming unusability: Developing efficient strategies in speech recognition systems". *Poster at Computer–Human Interaction 2000, ACM Conference on Human Factors in Computer Systems*, The Hague. http://www.crew.umich.edu/publications/00–02.pdf

Kim, D., Chan, H., Evermann, G., Gales, M., Mrva, D., Sim, K., and Woodland, P. 2005. "Development of the CU-HTK 2004 Broadcast News Transcription Systems". *ICASSP05*, Philadelphia.

Koester, H. 2004. "Usage, performance, and satisfaction outcomes for experienced users of automatic speech recognition". *Journal of Rehabilitation Research & Development* 41(5): 739–754. http://www.vard.org/jour/04/41/5/koester.html

Lambourne, A., Hewitt, J., Lyon, C., and Warren, S. 2004. "Speech-based real-time subtitling services". *International Journal of Speech Technology* 7: 269–279. http://homepages.feis.herts.ac.uk/~comrcml/IJSTarticle.pdf

Lincoln, M., Cox, S., and Ringland S. 1998. "A comparison of two unsupervised approaches to accent indentification". *International Conference on Spoken Language Processing*, Sydney.

Mangu, L., Brill, E., and Stolcke, A. 1999. "Finding consensus among words: lattice-based word error minimization". In *Proceedings of the 6th European Conference on Speech Communication and Technology*.

Nahamoo, D. 2006. Personal communication.

Pallett, D. 2003. "A look at NIST's benchmark ASR tests: Past, present, and future". http://www.nist.gov/speech/history/pdf/NIST_benchmark_ASRtests_2003.pdf

Saon, G., Povey, D., and Zweig, G. 2005. "Anatomy of an extremely fast LVCSR decoder". In *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon, 549–553.

Saraclar, M., Riley, M., Bocchieri, E., and Goffin. V. 2002. "Towards automatic closed captioning: Low latency real time broadcast news transcription". In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Denver.

Wald, M. 2005, "Personalised displays". In *Proceedings of the Conference on Speech Technologies: Captioning, Transcription and Beyond*. IBM. http://www.nynj.avios.org/Proceedings.htm
http://eprints.ecs.soton.ac.uk/11532/01/Personalized_Display_IBM_conference4.doc

Woodland, P., Pye, D., and Gales, M. 1996. "Iterative unsupervised adaptation using maximum likelihood linear regression". In *Proceedings ICSLP '96* www.asel.udel.edu/icslp/cdrom/vol2/754/a754.pdf.

CaptionMeNow. 2004. http://www-306.ibm.com/able/solution_offerings/captionmenow.html.

Author's addresses

Dimitri Kanevsky, Sara Basson, Alexander Faisman, Leonid Rachevsky, Alex Zlatsin
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598
USA

Sarah Conrod
P.O. Box 5300
1250 Grand Lake Rd
Sydney, NS, B1P 6L2 Canada
Primary Contacts:
kanevsky@us.ibm.com
sarah_conrod@capebretonu.ca

General Information requests can be sent to

lrachevs@us.ibm.com
zlatsin@us.ibm.com
sbasson@us.ibm.com
alexf@us.ibm.com

About the authors

Dimitri Kanevsky is a Research Staff Member at IBM Thomas J. Watson Research Center. He holds 90 patents and has a "Master Inventor" status. He is a finalist of the 2005 Word Technology Award for IT-Sofware. His conversational biometrics based security patent was recognized by MIT, Technology Review, as one of five most influential patents for 2003.

Sara Basson currently works in IBM Research, where she is developing accessibility opportunities for all users of technology, irrespective of their abilities or disabilities. She holds an M.B.A. from NYU, and a Ph.D. in Speech and Hearing Sciences from The Graduate Center of the City University of New York. She serves on the Board of Directors of AVIOS.

Alexander Faisman works at the IBM Thomas J. Watson Research Center. He developed ViaScribe software, which is the hub of the Liberated Learning project, a consortium of 10 universities worldwide. Alexander received a Masters of Science in Mathematics in 1985 from Tashkent University. He is also the author of many scientific papers presented in international conferences as well as in publications, and he holds numerous granted and pending patents.

Leonid Rachevsky is a software systems analyst in the "Human Language Technologies" group at the IBM T.J. Watson Research Center. He obtained his MSc in Mathematics at Kazan State University and his PhD in Technical Science (Applied Mathematics) at the Kazan Institute of Chemical Engineering. He has worked extensively as a software engineer and senior software analyst in Israel, Canada, and the United States.

Sarah Conrod is Lead Researcher of the Baddeck Showcase and Executive Director of the Alexander Graham Bell Institute at Cape Breton University. Her work centers on developing and implementing projects that blend technology and accessibility. She also sits on the Board of Directors for the Cape Breton Branch of the Canadian National Institute for the Blind. Sarah holds a Bachelor of Science, a Certificate in Public Administration, and a Masters of Education in IT.

Alex Zlatsin is a senior software engineer at IBM T.J. Watson Research Center. His latest projects include the development of the managed client systems for AIX and Linux and Grid computing, in particular the deployment of a Speech Recognition Engine as the voice grid application. He holds over 30 patents and is an IBM master inventor. He holds the MS degree in Applied Mathematics from Moscow University of Railway Transport.

Appendix 1: Adaptable user interface

Most UI experiments are designed as a sequential process: Identify the parameters one wants to test (such as: "is interface X better than interface Y"). Identify dependent variables: ("which interface leads to increased speed, and higher user satisfaction as determined by A,B,C…"). Conduct the experiment, analyze performance along the dimensions tested, such as "Customers preferred interfaces that allowed them to speak naturally, that allowed them to barge in, and that required less than 3 minutes to complete". That is incremental progress, however. Our suggestion is to come up with statistical methods to predict a priori which interfaces are most successful (fastest, most accurate, most fun) for users with a range of differences (language differences, learning preferences, etc.). In what follows we present a sample approach on how statistically based interfaces can be created.

We will consider a problem of creating the user editing interface that allows the user to correct decoding errors and that satisfies some conditions (e.g., requires relatively small human intervention, provides high user satisfaction, etc.). We consider a special case in which a user (editor) edits speech recognition errors by viewing a decoded text in a window. We can assume that a potential error W in a decoded script is identified using a confidence score $0 < C < 1$ (like in Mangu 1999). We assume also, for simplicity, that we have the following tools that can be used to correct the error:

T1. Display a list of alternative words L that are acoustically similar to the decoded word W (without playing audio). We associate a confidence score C(T_1, W) with the assumption that the correct word is located in the list of alternative words.
T2. Display an N-best list (Mangu 1999) of decoding sentences corresponding to W (without playing audio). We associate a confidence score C(T_2, W) with the assumption that the correct word is located in the N-best list.
T3. Display an orthographic phonetic representation corresponding to W. We associate a confidence score C(T_3, W) with the assumption that phonetic representation correctly matches W.
T4. Display a list of alternative words L corresponding to the decoded word W (by playing audio).
T5. Display an N-best list of decoding sentences corresponding to W (by playing audio).
T6. Listen to audio.

Next, we assume that these tools require different time for an editor to execute. Assume for simplicity, that the tool T1 is the fastest and the tool T6 is the most time consuming. But the tool T1 has a higher probability that the correct word is not chosen and T6 guarantees that the correct word will be chosen.

Now we will formulate the following statistical problem. We will model an editing interface as a sequence of observations O = o_1 o_2 … o_k associated with a displayed decoding text and produced by an unobservable "true" state sequence S = s_1 s_2 …s_k that is a correct (reference text) associated with a decoded text. Here each observation o_i is described as follows: o_i contains features that are observed on a display by a user (like a part of decoding text in a window, e.g., a decoded word W_j, and/or a confidence score C(T_1, W_j) of this word) and some features that are stored in a system memory, i.e., they are "observed" by the system (e.g., confidence scores C(T_i, W_j)). To each pair of states (o_i, s_i) we associate a tool t_i with the following property.

Application of the tool t_i to the observed state o_i generates a correct segment of the reference text s_i (e.g., a correct word) and an application of the tool t_i satisfies some set of conditions (e.g., the fastest among other tools that also generates the correct textual segment, or requires less attention etc.). In other words, the tool t_i is chosen if it satisfies some set of conditions and corrects the decoding error in o_i. Let us denote this sequence of tools t_1, t_2 … t_k associated with (o_1, s_1) (o_k, s_k) as T.

With this framework the goal of our user interface "recognition" system is to find a sequence T of interface tools with the maximum posterior probability given an observed sequence O. In other words, T' = argmax_{T} Prob{T, O}. In training data a set of typical (t_i, o_i, s_i) can be found that satisfies requirements (detect a correct word, small human intervention, high user satisfaction etc.). We can use direct modeling — Maximum Entropy (ME) — to estimate the conditional distribution Prob(T|O) using this training data as described in (Berger 1996). ME modeling is based on a principle of avoiding unnecessary assumptions. Training is regarded as supervised learning. The mathematical interpretation of the ME principle results in a constrained optimization problem: maximizing the entropy of the conditional distribution subject to given constraints. The direct modeling allows for potential sources of data in a unified fashion; for example, one can introduce a multitude of other features as user characteristics (age, attention, skills, nationality etc.) and environment characteristics (working in office, at home, in a car etc.), multiple editors, hierarchical tasks (e.g., speech and translation). One need also to investigate other approaches. For example, one could define a utility measureU({Ti, Tj, … }) as some combination of the probability that the sequence would result in a correct edit and the total required effort (which could be the sum of the times for each t_i).

To implement this algorithm, one would need to collect user interface data for different kinds of interfaces from many editors who would edit large amount of corpora (speech decoding, translations) using different kinds of user interfaces and different kinds of processes (re-speaking, automatic speech recognition, machine translation etc.).

Appendix 2: The algorithm for Content Spotting

a. We make a preliminary partition of a reference text into sentences. This can be achieved by inserting punctuation marks, or using punctuation that is already in place.
b. We introduce a notion of a *distance* between words and a *ball* around a word of some *radius*. For this we can use either timing or symbols in a text, like empty spaces. For example, let us consider the sentence: "Relatively many people attend the Alexander Graham Bell museum during one year". Using empty spaces in this sentence one can see that the distance between the word "attend" and the word "museum" will be exactly 5 empty spaces. There are the following words in the ball of radius 2 around the word "one": "museum", "during", "year".
c. We find in the text words that came with the input (decoded) sentence (that was produced by a speech recognizer).
d. We choose in the reference text words that are close to the words in the input sentence. The notion of how it is "close" we find empirically. For example, the first input word was discovered in the text at the 14th place, the second input word was discovered at the 16th place and also at the 113th and 268th places. In this case we choose a suitable segment in the text around the 14th place since the second input word has a distance from the first input word in the text equal 2 and other locations of these words are more distant.
e. If one discovers several such suitable segments in the reference text then the approximate search is done for all these segments and not for the whole reference text. In such

segments we can use comparison of sequences of words in the text and of the input phrase using Levenstein (string edit) distance.

This method gives good results for content spotting in real time. It is convenient for use with speech recognition decoding input sentences, since the reference text and the speech recognition input sentences consist of words that are written orthographically correctly.

Computer-aided translation as a distributed cognitive task

Barbara Dragsted
Copenhagen Business School

The present article examines the potential effects on the translation process of working interactively with a translation memory (TM) system, a tool for storing and sharing previous translations. A TM system automatically divides the source text into sentences presented to the translator one-by-one. Based on observations made in an empirical study of six professional translators and six translation students, it is argued that full sentences do not constitute a central cognitive processing category in translation, and that the sentence-by-sentence presentation inherent in TM systems therefore creates an unnaturally strong focus on the sentence, which affects the very task of translation (as well as the translation product). Particular attention is given to the impact of the use of TM systems on the informants' revision behaviour and their tendency to change the sentence structure.

Keywords: human–computer interaction, cognitive translation processes, computer-aided translation, translation memory systems, working memory

1. Introduction

The article, based on an empirical investigation carried out as part of a PhD project at the Copenhagen Business School in 2004 (Dragsted 2004, 2005), investigates collaborative cognition in the form of interaction between user (human translator) and artifact (a TM system) in translation and its effect on the way in which translators cognitively process the text being translated.

The focus is on whether translators let the TM system take control of their translation process and to what extent the system's design changes the cognitive task of translating. More specifically, the article will focus on changes in revision and editing behaviour and in the frequency with which translators change the sentence structure of the source language (Danish) when translating into the target language (English).

Pragmatics & Cognition 14:2 (2006), 443–464.
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

2. Theoretical background

2.1 Human–computer interaction in translation

Since the emergence of the first TM systems in the early 1990s, they have become one of the most widely used computer-aided translation tools by translation agencies and other organisations involved in translation. However, very little research has been done on how these systems affect the human translation process.

A TM system is a database which stores each source language (SL) and target language (TL) sentence pair as the translator progresses with the translation, and subsequently retrieves stored sentence pairs if the same or a similar sentence recurs in the same or another text. TM systems are used by individual translators to store and retrieve their own previous translations and ensure consistency, and by teams of translators within the same organisation or collaborating on the same or related texts. In both cases, translations are produced, in Norman's terminology, using "Cognition in the head and in the world" (Norman 1993).

The use of TM systems by groups of translators working towards a common goal, namely to produce a TL text on the basis of a SL text, may be considered an instance of distributed cognition, or more accurately, collaborative cognition (Harnard 2005), in the traditional sense that it is a joint activity involving several agents, both human and technological: Via a common database (the TM system), several translators share cognitive resources and build up a common understanding of a SL text as well as agreement (hopefully) on how the TL text should be formulated. So, rather than having the sole authority to decide the wording of the TL text, the individual translator 'surrenders' part of the decision to the TM system, and, via the TM system, to other translators in the group. The task of translating is distributed across the translator and other translators via an artifact (the TM system).

The individual translator may also work alone with the TM system to reuse his/her own previous translations, in which case cognition is simply distributed between user and artifact, the artifact being designed to remember large amounts of information for the translator. The present article deals with collaborative cognition mainly from a human–computer interaction (HCI) perspective (Wright et al. 2000; Norman 1986), investigating how the design of the TM system (the external representation) interacts with the translator's mental representation (internal representation) and the extent to which the system changes the translation task seen from the translator's perspective (Norman 1991: 21).

According to Zhang (1997: 180) "external representations are not simply inputs and stimuli to the internal mind; rather, they are so intrinsic to many cognitive tasks that they guide, constrain, and even determine cognitive behavior". Based on this assumption, it will be analysed and discussed to what extent the use of a TM system changes the cognitive task of translation because of the design of such systems. In conventional TM systems (cf. 3.1 below), the translator works with an interface that presents one sentence at a time and provides the option to write a translation of the sentence, or, if a previous version of the same or a similar sentence is retrieved from the system's database, to accept, revise or reject the sentence.

Hence, the TM system's interface (the external representation) makes a sentence-based presentation of the source text, which may not necessarily correspond with the translator's mental representation of the text. This can be expected to have an impact on certain behavioural aspects of the translation process (and product) and on professional translators' and students' perception of the translation task. To the extent that the informants feel that they process the text in a different way when they work with a TM system, and that using the TM system is "itself a task" (Norman 1991: 21), then the nature of the translation task has indeed changed, in accordance with Norman's personal view of an artifact (Norman 1991: 19ff).

2.2 Cognitive processes in translation

In order to make observations on the possible effect of a TM system on certain factors of the translation process, it was found necessary first to analyse and describe the same factors without the integration of the TM condition. Thus, the first step was to make observations on the 'pure' translation process with respect to the status of the sentence in cognitive processing in general and the informants' behaviour in terms of revision and changes to the sentence structure in particular.

The examination of the cognitive status of the sentence in the translation process was based on theories of cognitive psychology, notably theories of working memory (e.g., Miller 1967; Broadbent 1975; Crowder 1976; Baddeley 1986; Ericsson and Kintsch 1995), and of temporal aspects of text production, in particular pause theories (Schilperoord 1996; Butterworth 1980; Krings 1986a, 1986b). On the basis of these theories, the central cognitive category in translation, the 'translation unit', was defined as a segment that is limited by the capacity in working memory and identifiable through pauses in the translation process (Dragsted 2005).

3. Data and methods

3.1 Study design

The findings reported and discussed in this article were collected in a pilot study of six professional Danish-English translators with at least two years' experience and six final-year translation students at the Copenhagen Business School. The informants were assigned two different translation tasks. In the first task, observations were made on the 'normal' translation process without the use of a TM system, and the second task focussed on identifying possible effects of using a TM system. In both tasks, the twelve informants were assigned with translating a short text from Danish (the mother tongue) into English (the first foreign language).

Data from the first task were collected in *Translog*, a computer logging program, developed by Jakobsen in 1998, which allows the experimenter to record and analyse all keyboard and mouse activities carried out during the translation process, as well as time delays between keystrokes, i.e. pauses in the translation process (Jakobsen and Schou 1999). Based mainly on Schilperoord (1996), each pause was assumed to signal a boundary between two cognitive units in the translation process (Dragsted 2004: 65ff and 101ff). The data from *Translog* furthermore showed the amount of time spent on the actual production phase and on the revision phase respectively.

The second task involved the translation of a text similar to the one translated in the *Translog* experiment, but this time using a TM system, *Translator's Workbench 5.5* from TRADOS®. *Translator's Workbench* is a conventional TM system in the sense that it divides the source text into sentences and searches for identical or similar sentences in a sentence database. Although alternatives to the purely sentence-based working method are gaining ground (e.g., in the form of 'phrase banks' or bitext[1]), other TM systems currently available on the market are based on the same basic principle, (e.g., DejaVú from Atril, SDLX from SDL and Transit from STAR). Therefore, results similar to those reported in this article could be expected with other TM systems as well.

The data collected in the TM phase were not as detailed as the *Translog* data, and did not, for instance, provide information on pauses in the process.[2] The time spent on actual production and revision respectively was measured by asking the translators to indicate when they had finished translating the first draft ('when typing the last full stop').

The data collected from both tasks were subjected to statistical tests of significance and variance (ANOVA). Moreover, the quantitative data were supplemented with qualitative data in the form of the informants' retrospective verbalisations concerning the translation tasks. In the TM experiment, the data from the questionnaires constituted a direct data source in its own right, whereas in the *Translog* experiment, the informants' own observations were used mainly as support for the quantitative data.

3.2 Texts

The texts translated in both tasks were short business letters, containing 12 sentences totalling 212 words (*Translog* experiment) and 213 words (TM experiment) respectively.

The text translated in *Translog* was based on an authentic text, but company and personal names had been made anonymous to avoid recognition, and the text had been adapted in terms of sentence structure. The text did not contain complicated terminology and was described by all informants as a fairly straightforward translation task.

The text translated in the TM system was constructed on the basis of the text translated with *Translog*, i.e., the overall content, the level of difficulty, the number of words and the sentence structure were the same, which facilitated the comparison of the translation process in the two tasks, although, admittedly, this also implied a risk of a certain repetition effect. However, the two texts were translated with at least two weeks apart in order to reduce or even eliminate this effect.

In order to investigate the tendency of the informants to change the sentence structure when translating with the TM system, both texts had been constructed in a way that prompted such restructurings in that they contained a number of very short ('too' short) sentences and a number of very long ('too' long) sentences.

The results reported and discussed below thus apply to short texts of a commercial nature translated from Danish (the mother tongue) into English (the first foreign language) in an experimental setting. It should be stressed that the results might be different if one or more of these parameters were changed. Notably, language direction can be expected to have played an important role (Mogensen 2000), and the findings of this study cannot be generalized to other languages.

4. Results and discussion

4.1 The status of the sentence in the translation process

The question of which language segment constitutes the most central 'translation unit' has been a subject of discussion among translation scholars for many years. From a linguistic point of view, it has been argued that the translation unit is a flexible segment the size of which ranges from morpheme over word, group and clause to the sentence, the overall argument being that the translation unit is generally "as small as possible and as large as necessary" for establishing translational equivalence (e.g., Vinay and Darbelnet 1995 [1958]; Catford 1965; Newmark 1988). Other scholars have argued that the clause or the sentence is the basic translation unit, emphasising the importance of the proposition and the relationship between participant roles (e.g., Bell 1991; Hewson and Martin 1991), and finally, some scholars have argued that the translation unit is by necessity a long text segment ranging from a paragraph to a full text, since the structure and style of the whole text is more important than the content of individual words and sentences (e.g., Nida 1969; Bassnett-McGuire 1991).

In empirical studies of the translation process, it has been observed that the translation unit is typically a group or a clause containing 2–6 words (Gerloff 1986; Lörscher 1986; Krings 1986b; Kiraly 1995).

In the present study it was similarly found that under normal circumstances (without the TM condition) the informants all appeared to work with either clauses or phrases as the most frequent translation unit:[3] Four out of six professional translators most frequently processed clauses as the translation unit, and two out of six most often worked at phrase level. In the group of students, five out of six had most segments at phrase level and one student most frequently worked with clauses.

A full sentence was not a frequent segment type in either of the groups and only accounted for approx. 2.5% on average of the cognitive segments processed by the informants in both groups (in comparison, clauses and phrases each accounted for 30–40% in the two groups). This low frequency of sentence segments was expected considering the sometimes complex nature and large number of words in a sentence and the limited capacity in working memory. Since the text contained a number of very short, non-complex sentences with only one proposition, which were therefore essentially no different from clauses, most of the informants had at least one full sentence segment.

Segments going across two sentences were found with a frequency of 5% on average in the group of professionals, but only 0.5% in the group of students, i.e., there was a certain tendency to work across sentence boundaries (i.e. without making a pause between two sentences) in the group of professional translators that was not found in the group of students. This tendency was confirmed by the informants' own observations: Only one of the six professional translators stated that she usually read (and therefore paused before) a new sentence before she translated it, whereas the others apparently went ahead with the translation without pausing before each new sentence. In the group of students, five of the six informants clearly stated that they read an entire sentence before starting to translate it.

Generally, the *Translog* data and the retrospective verbalisations did not offer support for the sentence being a central cognitive category in the translation process. In neither of the groups did it constitute a frequent translation unit. In particular the professional translators did not, apparently, regard the sentence as being central to their processing, whereas the students most often made a pause before commencing the translation of a new sentence, which indicated that sentence boundaries were more relevant in their cognitive processing. Based on these observations and conclusions, experiments were carried out with *Translator's Workbench* to investigate the effect in the two groups of integrating a TM system, with its sentence segmentation, into the translation task.

4.2 TM systems' effect on the translation task

The integration of the TM system was found to affect the cognitive task of translation mainly in two ways (as observable in the quantitative data): The way in which the informants behaved with respect to revision of the translation and their tendency to change the sentence structure of the source text when producing the target text.

4.2.1 *Effect on revision*

It was found in the *Translog* experiment (without TM) that the professional translators spent considerably more time than the students on revising the text after having produced a draft translation, cf. the '*Revision time without TM*' column in Table 1. One possible explanation for this difference could be that professional translators, to a greater extent than students, worked quickly through the text in order to build up a general understanding of the text. When making this first draft, the professional translators did not make long pauses to consider, for instance, lexical problems. Such difficulties were solved in the revision phase when the translators had 'the complete picture' of the text. In other words, many of the important decisions were postponed until the end.

The students, on the other hand, tended to deal with lexical and other problems as they arose, and generally seemed to work more thoroughly through each sentence as they translated the text. This seems to be a plausible explanation for the somewhat shorter revision phase in the group of students. Still, all the students did spend some time revising their translations at the end in the *Translog* experiment (see Table 1).

The introduction of a TM system into the translation process changed the informants' revision behaviour. The figures in Table 1 show the effect of the TM system on time spent on revision.

Table 1. Revision time with and without the use of TM

Professionals[4]	Revision time without TM[5]	Revision time with TM	Revision time effect
PT1	183 sec. (18%)	115 sec. (13%)	−68 sec. (−37%)
PT2	472 sec. (43%)	171 sec. (27%)	−301 sec. (−64%)
PT3	195 sec. (29%)	194 sec. (25%)	−1 sec. (−0.5%)
PT4	356 sec. (49%)	310 sec. (45%)	−46 sec. (−13%)
PT5	364 sec. (32%)	171 sec. (14%)	−193 sec. (−53%)
PT6	152 sec. (20%)	0 sec. (0%)	−152 sec. (−100%)
Average	287 sec. (32%)	189 sec. (26%)	−98 sec. (−34%)

Students	Revision time without TM	Revision time with TM	Revision time effect
ST1	99 sec. (9%)	0 sec. (0%)	−99 sec. (−100%)
ST2	166 sec. (18%)	120 sec. (13%)	−46 sec. (−28%)
ST3	772 sec. (49%)	0 sec. (0%)	−772 sec. (−100%)
ST4	207 sec. (12%)	0 sec. (0%)	−207 sec. (−100%)
ST5	165 sec. (7%)	116 sec. (7%)	−49 sec. (−30%)
ST6	52 sec. (4%)	0 sec. (0%)	−52 sec. (−100%)
Average	244 sec. (17%)	40 sec. (3%)	−204 sec. (−84%)

All informants in both groups spent less time on final revision with the TM system than they did without it, yielding a significant difference as regards revision time across the groups ($F(1,10) = 7.039$, $p = .0242$).

The effect was particularly marked in the group of students where, on average, the revision time was reduced by 84%, and four out of six students, [ST1], [ST3], [ST4], and [ST6], left out the final revision phase altogether. ANOVA showed a significant reduction in revision time in the group of students when they used a TM system (F(1,10) = 5.356, p = .043).

In the group of professionals, the reduction in revision time was not as substantial as in the group of students (34%), and five out of six professionals still spent some time on final revision of the target text. ANOVA testing showed no significant difference in the group of professionals when comparing the *Translog* task with the TM task. Nevertheless, the tendency to spend less time on revision, also in the group of professionals, seems to be unmistakable.

Arguably, as mentioned in Section 3.2, the similarity of the two texts may have involved a certain repetition effect meaning that the reduced revision time found with the TM system merely reflected the informants' familiarity with the text. However, the two-week time span renders this explanation implausible; even if the informant had translated a similar text before, after two weeks the recollection of the text would be so 'blurred' that the informant would normally feel the same need to check his/her translation. The fact that the texts did not contain complicated terminology offers further support for this argument, since the informants could not be expected to save time the second time around on account of time-consuming terminology research carried out the first time.[6]

Hence, it seems more likely that the tendency in both groups, and mainly among the students, to spend less time on revising the target text after having produced a draft translation does not imply that the informants did not revise their translations as thoroughly as before, but rather that the revision behaviour appears to have changed. The reduced time spent on revision after producing the translation arguably reflects an increase in revision time during the production phase: Part of the time the informants spent on revising the target text after the actual production of the text without the TM, they now spend on revising one sentence at a time as they go along (with the TM).

Further support for the changed revision behaviour can be found when relating the figures for time spent on revision in the group of professionals to their total time figures (see Table 2).

Most of the informants in the group of professionals spent either more time or approx. the same time on the total translation task under the TM condition than they did in the *Translog* setting. At the same time, most of them spent less time on end revision. This suggests that most of the professionals spent more time on the actual production task with the TM system, which again suggests, although it does not prove, that the professional translators spent more time on refining each individual sentence when they worked with a TM system.

Table 2. Total time with and without the use of TM

Professionals	Total time without TM	Total time with TM	Total time effect
PT1	17min. 22sec. (1042 sec.)	14min. 12 sec. (852 sec.)	−190 sec. (−18%)
PT2	18min.10sec. (1090 sec.)	10min. 43 sec. (643 sec.)	−447 sec. (−41%)
PT3	11min. 10sec. (670 sec.)	12min. 45sec. (765 sec.)	+95 sec. (+14%)
PT4	12min. 07sec. (727 sec.)	11min. 25sec. (685 sec.)	−42 sec. (−6%)
PT5	18min. 56sec. (1136 sec.)	19min. 49sec. (1189 sec.)	+53 sec. (+5%)
PT6	12min. 54sec. (774 sec.)	19min. 50sec. (1190 sec.)	+416 sec. (+54%)
Average	907 sec.	887 sec.	−20 sec. (−2%)

Students	Total time without TM	Total time with TM	Total time effect
ST1	19min. 08sec. (1148 sec.)	12min. 48sec. (768 sec.)	−380 sec. (−33%)
ST2	15min. 17sec. (917 sec.)	15min. 46sec. (946 sec.)	+29 sec. (+3%)
ST3	26min. 10sec. (1570 sec.)	16min. 46sec. (1006 sec.)	−564 sec. (−36%)
ST4	29min. 26sec. (1766 sec.)	29min. 20sec. (1760 sec.)	−6 sec. (−0.3%)
ST5	37min. 06sec. (2226 sec.)	29min. 23sec. (1763 sec.)	−463 sec. (−21%)
ST6	20min. 21sec. (1221 sec.)	19min. 30sec. (1170 sec.)	−51 sec. (−4%)
Average	1475 sec.	1236 sec.	−239 sec. (−16%)

The apparent change in revision behaviour in both groups triggers the question of whether the informants observed this change. It appears from the questionnaires that a majority of informants in both groups found that the translation process was generally affected by the integration of a TM system with its sentence segmentation, but only one informant, [PT4], specifically mentioned the influence of the system on the time she spent on revision: *"I spend less time on proof-reading and revision after working with Trados, probably because I tend to work more thoroughly on each sentence from the beginning"*. This statement seems to reflect what most of the informants did, but only one informant actually seemed to be (fully) aware of this.

Even though the actual difference in revision time was substantially more marked in the group of students, it might be argued that the behavioural difference was equally important in the group of professionals. As shown above, one of the ways in which the professionals differed from the students in the *Translog* setting was that they spent relatively more time on final revision, in fact almost twice as much time relative to total production time (32% on average in the group of professionals vs. 17% in the group of students, cf. Table 1). So, whereas in the *Translog* setting the professionals postponed many decisions to the end revision, in the TM setting suggestions were often evaluated instantly, which implied a change in the professionals' behaviour and the translation task.

Since only one informant seemed to be aware of this change, the integration of a TM system involves the risk that the professionals, without being consciously aware of this, may not to the same extent have ensured the coherence of the target text. The same applies to the students, who had a significant revision time reduction, but in this group, the final revision phase was relatively less important even without the TM condition, and consequently the relative behavioural effect was less marked than in the group of professionals. For a further discussion of the risk of incoherence as a result of doing sentence-by-sentence translation see Bédard (2000), who cautions about the translated text proving to be 'a "sentence salad" of questionable taste'.

In conclusion, it may be argued that a side effect of the interaction between translator and TM system is a behavioural change with respect to revision, which may be explained by the TM system's increased focus on sentences. Rather than revising the target text in its entirety at the end, the informants tended to revise each sentence more thoroughly when they used a TM system, demonstrating how the TM system's interface (external representation) seems to have affected the HCI strategy adopted by users (Wright et al. 2000: 34; Zhang 1997: 180).

The changed revision behaviour may be seen as a case of the informants *unconsciously* handing over control to the TM system (as opposed to what will be discussed in the next section), i.e., even though changes in translation behaviour may actually be observed, the informants are not (with one exception) aware of these behavioural changes. Arguably, from the point of view of the translator interacting with the TM system, the system does not change the translation task with respect to revision (Norman 1991: 20).

4.2.2 *Effect on changes to sentence structure*

As mentioned in 3.2 above, the texts translated in both tasks had been adapted in a way that prompted restructuring of the sentence structure in the SL text. Consequently, in the *Translog* setting (without TM) the professionals changed the sentence structure 3 times on average (see Table 3), a frequency that can be considered rather high given the length of the text (12 sentences). The students, on the other hand, had very few instances of sentence structure changes, *viz.* 0.5 on average, reflecting the fact that only one informant in the group of students, [ST4], deviated from the SL sentence structure in her translation into the TL (3 times) (see Table 3).

One of the hypotheses to be evaluated in the present article was that the informants would change their behaviour in terms of sentence structure alterations, reflecting a change in their mental representation and the way they

Table 3. Instances of sentences structure alterations with and without the use of TM

Professional	Sentence structure alterations without TM	Sentence structure alterations with TM	Sentence structure alteration effect
PT1	2	–	÷2 (÷100%)
PT2	4	2	÷2 (÷50%)
PT3	3	3	–
PT4	2	1	÷1 (÷50%)
PT5	3	4	+1 (+33%)
PT6	4	3	÷1 (÷25%)
Average	3	2.2	÷0.8 (÷27%)

Student	Sentence structure alterations without TM	Sentence structure alterations with TM	Sentence structure alteration effect
ST1	–	1	+1 (+100%)
ST2	–	–	–
ST3	–	–	–
ST4	3	–	÷3 (÷100%)
ST5	–	–	–
ST6	–	–	–
Average	0.5	0.2	÷0.3 (÷60%)

processed the text in general when they worked with a TM system. The figures relevant for evaluating this hypothesis are presented in Table 3.

The figures in Table 3 reveal a slight tendency among the informants in both groups to have more sentence structure alterations when they did not use a TM, although the differences are not as convincing as might have been expected.

On average, however, the informants in both groups were less inclined to change the structure of the source text under the TM condition, and a closer look at the individual informants in the group of professionals showed that four out of six were less prone to combine or split up source text sentences when they used a TM. In the group of students, the individual analysis does not seem to make much sense because of the very low number of informants who made sentence structure alterations at all, but as appears from the table, the one informant, [ST4], who did indeed have quite a considerable number of structural changes in the *Translog* setting, made no such changes with the TM.

The decrease in the number of sentence structure alterations in the TM setting may be explained by different factors. One possible explanation is that the TM system's interface made the informants become so focused on each individual sentence that they disregarded the surrounding context and consequently did not even consider changing the structure of the source text.

Another possible explanation is that the informants were aware that the translation product might gain from changing the sentence structure of the source text, but the technical 'surplus cost' of making such changes made them less inclined to perform the additional actions required to combine or split up sentences.[7]

Finally, a third explanation could be that under normal circumstances (without a TM), changes to the sentence structure would be postponed to the final revision phase during which the translator reads through and revises the target text in its entirety. When, as described above, the importance of end revision was reduced in the TM setting, this may have had the implication that sentence structure alterations that would normally have been made during the revision phase were omitted. Consequently, there may be a relationship between the decreased number of changes to the SL sentence structure and the time spent on end revision. For technical reasons[8] it moreover became difficult to change the structure of sentences after completing the actual production, which further complicated, and potentially reduced the number of, structural changes.

The informants' observations regarding the effects of sentence segmentation were analysed to throw more light on the weight of these possible explanations. It appeared from the questionnaires that most informants were aware of the impact of sentence-by-sentence presentation on the translation process. However, the two groups differed considerably with regard to the way they perceived the sentence segmentation carried out by the TM system. Table 4 below seeks to provide an overview of the informants' positive and negative comments on sentence segmentation.[9]

As appears from Table 4, the professional translators generally found the sentence-by-sentence presentation to be a disadvantage, because it made them more inclined to preserve the structure of the source text, e.g., [PT3], [PT5], [PT6]. Only one informant in the group of professionals, [PT1], stated that she thought it was an advantage that the TM system divided the text into sentences.

In the group of students, on the other hand, none of the informants found the sentence segmentation to be an absolute disadvantage, and four out of six, [ST1], [ST2], [ST3], and [ST4], clearly considered the division into sentences to be an advantage, precisely because it enabled them to concentrate on one sentence at a time.

The fact that most of the students considered the sentence-by-sentence interface to be helpful suggests that they were not, apparently, aware of the potential problems related to the sentence presentation and consequently did not consider changing the sentence structure of the source text (cf. the first

Table 4. Observations on sentence segmentation (advantages vs. disadvantages)

Professional	Advantages	Disadvantages	Conclusion
PT1	"[…] nice to see each individual sentence in isolation"	"[…] start translating right away without gaining an overview of the text"	ADVANTAGE
PT2	"TWB is suitable for large assignments, particularly in the domain of IT, because they are structured in a way which is good to maintain"	"[…] requires a disproportionately large effort to achieve a good sentence structure"	DISADVANTAGE
PT3	"[…] it makes you focus on one sentence, which can also be an advantage in some cases, I guess"	"I see it more as a disadvantage than as an advantage. […] If the text […] contains sentences which are either too short or too long, it may be a problem to work with sentence segments"	DISADVANTAGE
PT4	"I spend less time on proof-reading and revision […], probably because you tend to work more thoroughly on each sentence from the beginning"	"A disadvantage […] You feel restricted by the order of sentences […]The fact that each sentence is presented in its own coloured box implies that you focus a lot on the individual sentence"	DISADVANTAGE
PT5		"[…] the text presented sentence by sentence. When you are not used to this, it restricts you. I really had to make an effort not to stick to the sentence boundaries"	DISADVANTAGE
PT6		"[…] It can get very abrupt, even though it is possible to combine sentences. But I can imagine that I wouldn't always do it, because it requires more planning. […] you tend to stick to the sentence structure in the source text"	DISADVANTAGE

possible explanation mentioned above). The professionals, on the other hand, were generally aware of the potential 'danger' related to sentence segmentation, but technical and planning-related issues apparently made them less inclined

Student	Advantages	Disadvantages	Conclusion
ST1	"[…] makes it easier not to lose track of what you are doing"		ADVANTAGE
ST2	"There is a lot of focus on sentences, and I think that is an advantage"		ADVANTAGE
ST3	"[…] nice to have the text presented sentence by sentence. It enables you to concentrate on one sentence at a time. It becomes easier to keep track of the text"	"[…] You have to keep an eye on the screen all the time to make sure that you translate the right piece of the text"	ADVANTAGE
ST4	"I don't think it is a disadvantage"	"I am not sure if it is an advantage […] perhaps it is a very mechanical way to translate, because you only see one sentence at a time"	MIXED
ST5	"I think it was an advantage because it made it easier to keep track of the text"		ADVANTAGE
ST6	"You can concentrate more on each individual sentence"	"[…] you lack the rest of the context"	MIXED

to change the sentence structure, also during the revision phase (cf. the second and third possible explanations above).

The difference between the two groups of informants as regards their view on the sentence-by-sentence presentation under the TM condition should be seen in the light of their processing under normal circumstances (i.e., without the TM): In the group of students, there were very few sentence structure alterations in the *Translog* setting as well as the TM setting. In general, the students rarely change the sentence structure of the source text, whether or not they work with a TM system, but rather they seem to follow a pattern characteristic of novices, rigidly adhering to guidelines and plans and without making additional judgments (Dreyfus et al. 1986). The students, strictly speaking, did not even consider it an option to make changes to the sentence structure, i.e. they were not aware that they could potentially improve the target text by changing the structure of the source text. Overall, the sentence-by-sentence interface seems to fit the students' normal cognitive behaviour and mental representation of the text, and they do not feel that the TM system fundamentally changes the translation task.

For most professional translators, however, it seems to be the case that the TM system's interface has a constraining effect on their cognitive behaviour and mental representation of the text and thus changes the translation task (Zhang 1997: 180; Norman 1991: 20). Several of the professional translators normally behave in a way typical of expert performers, i.e., they make ad hoc rather than rule-based decisions (Dreyfus et al. 1986) and are able to process large chunks of information, which, in the case of translation, would enable them to go beyond the boundaries of sentences (Ericsson and Kintsch 1995). Consequently, in the *Translog* setting, the professional translators quite often changed the source text sentence structure. When they worked with a system which impeded this strategy by focussing explicitly on one sentence at a time, the professionals were aware of this, and, in most cases, considered it to be a disadvantage, cf. Table 4, exactly because they felt that the system forced them into preserving the sentence structure of the source text. The acknowledgement of the potential risk of TM systems is one step towards avoiding it, and as shown in Table 3 above, most professional translators still had a fair amount of sentence structure alterations when they used a TM.

However, as also shown in Table 3, structural changes did not occur with the same frequency, which suggests that most professional translators were still influenced by the sentence-by-sentence presentation, even though they were aware of its potential weakness. As suggested above and as shown in Table 4, the professionals felt restricted by technical and planning factors. For instance, [PT2] stated that she thought it "[...] *requires a disproportionately large effort to achieve a good sentence structure*", and [PT6] found that "[...] *It can get very abrupt, even though it is possible to combine sentences. But I can imagine that I wouldn't always do it, because it requires more planning*".

Interestingly, the one informant in the group who considered the sentence segmentation to be an advantage, and apparently was not aware of the potential problems related to the TM design, [PT1], was also the only informant who made no sentence structure alterations at all in the TM setting (as opposed to two alterations in the *Translog* setting), i.e. she was the only professional translator who uncritically preserved the source text sentence structure.

Another experiment carried out in continuation of the one reported here (Dragsted 2004), where the informants were tasked with translating another text in the TM system, but this time divided into full paragraphs, showed that most of the subjects in the group of professionals strongly preferred paragraph segmentation over sentence segmentation as also found by Esselink (2000: 363). Bédard argues that even at a paragraph level, translators may well come to see only a collection of sentences and miss the flow of ideas (Bédard 2000).

Still, it seems safe to assume that many of the problems related to sentence translation will be overcome in a system operating at paragraph level.

The analysis of the informants' observations related to their actual behaviour as observed in the target text leads to the conclusion that awareness vs. unawareness was a decisive factor as regards the number of sentence structure alterations performed by the informants as well as the extent to which informants found it an advantage or a disadvantage to work with a TM system. Figure 1 below seeks to summarise this:

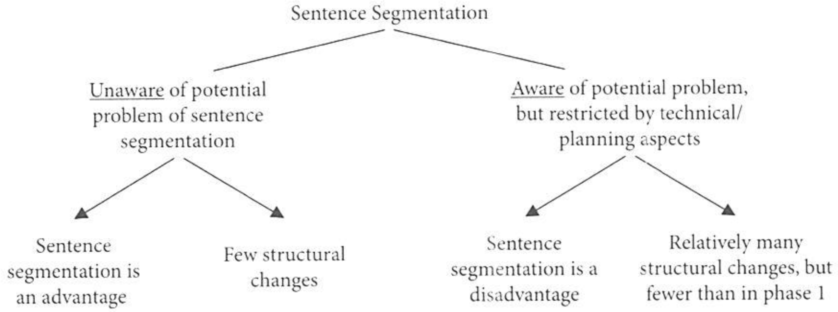

Sentence Segmentation
Unaware of potential problem of sentence segmentation
Aware of potential problem, but restricted by technical/ planning aspects
Sentence segmentation is an advantage
Few structural changes
Sentence segmentation is a disadvantage
Relatively many structural changes, but fewer than in phase 1

Figure 1. Impact of awareness on structural changes and informants' views

The professional translators were usually aware of the potential (negative) effect of the sentence-by-sentence interface. It might be argued, therefore, that the professional translators are *consciously* guided by the sentence structure suggested by the computer when they work with a TM system and feel that the system changes the translation task. The sentence-by-sentence interface has a determining effect not only on the physical, but also on the translators' mental representation of the text, which, strictly speaking, changes from one coherent text into a collection of sentences.

For informants (mainly students) who were unaware of the potential problem caused by the sentence-by-sentence interface, the impact (to the extent that there was an impact at all compared with the *Translog* setting) on strategic behaviour was unconscious, i.e. decisions on sentence structure were *unconsciously* left up to the TM system. In this case, the interaction with the TM system was not to the same extent considered to change the cognitive task of translating the text.

5. Conclusion

The purpose of the study presented in this article was to identify and discuss the possible effects on certain aspects of the translation process of working with a TM system, based on the hypothesis that the sentence segmentation performed by TM systems creates an unnaturally strong focus on sentences. In particular, it has been discussed to what extent the integration of a TM system has an effect on professional and non-professional translators' revision behaviour and on their tendency to make changes to the sentence structure of the SL text, reflecting a change in the way they process the text in general. It has been found that both of these two aspects were indeed affected by the TM condition. The informants' interaction with a TM system has thus been shown to guide and constrain — consciously or unconsciously — users' behaviour and to some extent change the nature of the cognitive task of translating a text seen from the translator's perspective.

Quantitatively, revision behaviour was affected in both groups, but mainly in the group of students, although it might be argued that the actual effect was equally or even more important in the group of professionals because, under normal circumstances, the professionals spent relatively more time on revising the text in its entirety at the end than the students did. Generally, neither of the groups was aware of the change in revision behaviour.

The number of changes to the sentence structure was slightly reduced in both groups, but mainly in the group of professionals, where four out of six made fewer structural changes with the TM system than without it. The professionals were aware of the problem related to sentence segmentation, and normally saw it as a disadvantage. They acted on the problem to a certain extent, resulting in a considerably higher number of changes than in the group of students. The students generally saw no potential problem in the sentence segmentation performed by TM systems. They consequently found it to be an advantage, and only changed the sentence structure in very few cases.

The informants' observations generally indicate that the professionals, who expressed strong reservations about the sentence segmentation in TM systems, found that the TM system changes the cognitive task of translation (in an inappropriate way), whereas the students did not to the same extent feel this change, but uncritically let themselves be guided by the sentence-by-sentence presentation.

These observations, together with the quantitative measurements, suggest a need for an adjustment of existing TM systems, where ideally focus is moved away from the sentence, and the text is divided into full paragraphs. Support for this suggestion was found in an experiment (with paragraph instead of sentence segmentation) carried out in continuation of the one reported here. As long as such a system is not available, novice translators (including students) should use TM systems with caution and be explicitly made aware of the potential problems of the system's design.

Notes

1. Bitext is a merging of a source text and its translation, which can be consulted using a full-text search tool. A bitext, as opposed to a TM, is designed for consultation by the user, not a machine, which means expectations for recycling sentences are less stringent (Arrouart and Bédard 2001). The possibility of automatically recycling fragments of sentences stored in a 'phrase bank' has been investigated for instance by McTait et al. (1999), McTait and Trujillo (1999) and Langé et al. (1997).

2. For technical reasons, it was not feasible to integrate *Translog* with the TM system.

3. As indicated previously, a 'translation unit' is defined here as a language segment that can be processed in working memory, and which is therefore limited in size. The size and nature of the translation unit may be identified on the basis of pauses in the flow of production, and may vary according to the level of expertise of the translator and the level of difficulty perceived (Dragsted 2005).

4. The abbreviations PT1–PT6 refer to professional translators, whereas the abbreviations ST1–ST6 refer to student translators.

5. The percentages in parentheses indicate the percentage of revision time of total production time.

6. A way to eliminate the possibility of the repetition effect completely would have been to change the order of the tasks for some of the informants. However, for practical reasons this could not be done, as the data from *Translog* were needed to construct (part of) the experiment with *Translator's Workbench* (cf. Dragsted 2004).

7. In order to change the sentence structure of the source text, the translator has to manually select from a drop-down menu that he/she wants to expand or shrink the segment.

8. The TM system adds a code at the end and beginning of each sentence, which cannot be deleted. Therefore, changes to the sentence structure, which frequently involve deleting full stops and the first letter in a new sentence (in the case of combining two sentences), cannot be performed without reopening the translation segment, which complicates the process.

9. The table only includes observations concerning sentence segmentation, i.e. comments regarding for instance technical features and time consumption have been left out.

References

Arrouart, C. and Bédard, C. 2001. "In praise of bitext". Retrieved 21 February 2006 from www.terminotix.com. Based on Arrouart, C. and Bédard, C. 2001. "Éloge du bitexte". *Circuit* 73: 30.

Baddeley, A. 1986. *Working Memory*. Oxford: Clarendon Press.

Bassnett-McGuire, S. 1991. *Translation Studies*. Revised edition. London: Routledge.

Bédard, C. 2000. "Translation memory seeks sentence translator...". Retrieved 21 February 2006 from www.terminotix.com. Based on Bédard, C. 2000. "Mémoire de traduction cherche traducteur de phrases...". *Traduire* 186: 41–49.

Bell, R.T. 1991. *Translation and Translating: Theory and Practice*. London: Longman.

Broadbent, D.E. 1975. "The magic number seven after fifteen years". In A. Kennedy and A. Wilkes (eds), *Studies in Long Term Memory*. London: John Wiley & Sons, 3–18.

Butterworth, B. 1980. "Evidence from pauses in speech". In B. Butterworth (ed), *Language Production. Vol. 1: Speech and Talk*. London: Academic Press, 155–176.

Catford, J.C. 1965. *A linguistic Theory of Translation*. London: Oxford University Press.

Crowder, R.G. 1976. *Principles of Learning and Memory*. Hillsdale, NJ: Lawrence Erlbaum.

Dragsted, B. 2004. *Segmentation in Translation and Translation Memory Systems — An Empirical Investigation of Cognitive Segmentation and Effects of Integrating a TM System into the Translation Process*. Copenhagen: Samfundslitteratur.

Dragsted, B. 2005. "Segmentation in translation: Differences across levels of expertise and difficulty". *Target* 17:1, 49–70.

Dreyfus, H.L. Dreyfus, S.E., and Athanasiou, T. 1986. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press

Ericsson, K.A. and Kintsch, W. 1995. "Long-term working memory". *Psychological Review* 102(2): 211–245.

Esselink, B. 2000. *A Practical Guide to Localization*. Amsterdam: John Benjamins.

Gerloff, P. 1986. "Second language learners. Reports on the interpretive process: Talk-aloud protocols of translation". In J. House and S. Blum-Kulka (eds), *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, 243–262.

Harnad, S. 2005. "Distributed processes, distributed cognizers, and collaborative cognition". *Pragmatics & Cognition* 13(3): 501–514.

Hewson, L. and Martin, J. 1991. *Redefining Translation. The Variational Approach*. London: Routledge.

House, J. and Blum-Kulka, S. (eds). 1986. *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr.

Jakobsen, A.L. and Schou, L. 1999. "Translog documentation". In *Probing the Process in Translation: Methods and Results. Copenhagen Studies in Language* 24. Copenhagen: Samfundslitteratur, 151–186.

Kiraly, D.C. 1995. *Pathways to Translation. Pedagogy and Process*. Kent: Kent State University Press.

Krings, H.P. 1986a. *Was in den Köpfen von Übersetzern vorgeht*. Tübingen: Gunter Narr.

Krings, H.P. 1986b. "Translation problems and translation strategies of advanced German learners of French (L2)". In J. House and S. Blum-Kulka (eds), *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, 263–276.

Langé, J-M., Gaussier, E., and Daille, B. 1997. "Bricks and Skeletons: Some ideas for the near future of MAHT". *Machine Translation* 12: 39–51.

Lörscher, W. 1986. "Linguistic aspects of translation processes: Towards an analysis of translation performance". In J. House and S. Blum-Kulka (eds), *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, 277–292.

McTait, K. and Trujillo, A. 1999. "A language-neutral sparse-data algorithm for extracting translation patterns". In *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation, TMI-99*, Chester, England, 98–108.

McTait, K., Olohan, M., and Trujillo, A. 1999. "A building blocks approach to translation memory". In *Proceedings of the 21st ASLIB International Conference on Translating and the Computer*, London.

Miller, G.A. 1967. *The Psychology of Communication. Seven Essays*. New York and London: Basic Books.

Mogensen, E. 2000. "Orwellian linguistics: How using computer-aided translation tools impacts the target language". *Language International*, October 2000, 28–31.

Newmark, P. 1988. *A Textbook of Translation*. London: Prentice Hall.

Nida, E. 1969. *The Theory and Practice of Translation*. Leiden: United Bible Societies.

Norman, D.A. 1986. "Cognitive engineering". In D.A. Norman and S.W. Draper (eds), *User Centered System Design. New Perspectives on Human–Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum, 31–61.

Norman, D.A. 1991. "Cognitive artifacts". In J.M. Carrroll (ed), *Designing Interaction. Psychology at the Human–Computer Interface*. Cambridge: Cambridge University Press, 17–38

Norman D.A. 1993. "Cognition in the head and in the world: An introduction to the Special Issue on Situated Action". *Cognitive Science* 17: 1–6.

Schilperoord, J. 1996. *It's about time: Temporal Aspects of Cognitive Processes in Text Production*. Amsterdam: Rodopi.

Vinay, J.P. and Darbelnet, J. 1995. *Comparative Stylistics of French and English. A methodology for translation*. Translated and edited by J.C. Sager and M.-J. Hamel. Amsterdam/Philadelphia: John Benjamins.

Wright, P.C., Fields, R.E., and Harrison, M.D. 2000. "Analyzing human–computer interaction as distributed dognition: The resources model". In *Human–Computer Interaction, volume 15*. Hillsdale, NJ: Lawrence Erlbaum, 1–41.

Zhang, J. 1997. "The nature of external representations in problem solving". *Cognitive Science* 21(2): 179–217.

Author's address

Barbara Dragsted
Department of English, CRITT
Copenhagen Business School
Dalgas Have 15
DK-2000 Frederiksberg
Denmark

bd.eng@cbs.dk

About the author

Barbara Dragsted has a PhD in translation from the Faculty of Languages, Communication and Culture at the Copenhagen Business School, where she is currently employed as an assistant professor teaching translation and English grammar. She is involved in different research projects under the Center for Research and Innovation in Translation and Translation Technology (CRITT). Her research interests include cognitive processes in translation, translation technology, terminology and knowledge sharing, and the study of translation workflows in multilingual communication.

Access to on-line full text

John Benjamins Publishing Company's journals are available in on-line full-text format as of the volume published in 2000. Some of our journals have additional (multi-media) information available that is referred to in the articles.

Access to the electronic edition of a volume is included in your subscription. We offer a pay-per-view service per article for those journals and volumes to which you did not subscribe.

Full text is provided in PDF. In order to read these documents you will need Adobe Acrobat Reader, which is freely available from **www.adobe.com/products/acrobat/readstep2.html**

You can access the electronic edition through the gateways of major subscription agents (SwetsWise, EBSCO EJS, Maruzen) or directly through IngentaConnect.

If you currently use **www.ingenta.com** or **www.ingentaselect.com** (formely, Catchword) to access your subscriptions, these rights have been carried over to **www.ingentaconnect.com**, the new, fully merged service. All bookmarked pages will also be diverted to the relevant pages on **www.ingentaconnect.com**.

If you have not yet set up access to the electronic version of the journal at IngentaConnect, please follow these instructions:

If you are a personal subscriber:

- Register free at **www.ingentaconnect.com**. This is a one-time process, that provides IngentaConnect with the information they need to be able to match your data with the subscription data provide by the publisher. Your registration also allows you to use the e-mail alerting services.
- Select *Personal subscriptions.*
- Select the publication title and enter your subscription number. Your subscription number can be found on the shipping label with the print journal, and on the invoice/renewal invitation.
- You will be notified by email once your online access has been activated.

If you are an institutional subscriber:

- Register free at **www.ingentaconnect.com** by selecting the registration link and following the link to institutional registration.
- Select *Set up subscriptions.*
- Select the publication title and enter your subscription number. Your subscription number can be found on the shipping label with the print journal, and on the invoice/renewal invitation.
- You will be notified by email once your online access has been activated.

If you purchase subscriptions via a subscription agent they will be able to set up subscriptions on IngentaConnect on your behalf – simply pass them your IngentaConnect ID, sent to you at registration.

If you would like further information or assistance with your registration, please contact **help@ingentaconnect.com**.

For information on our journals, please visit **www.benjamins.com**

Guidelines for Contributors

1. Contributions should be in British or American English consistently. If not written by a native speaker of English the paper should be checked by a competent native speaker.
2. All manuscripts should be accompanied by a biographical note (50-75 words), an abstract (100-150 words), a short list of keywords and the author's name, affiliation, contact address, email and home page (if available).
3. MANUSCRIPTS can be submitted either by regular mail (in triplo, double spaced with wide margins) or electronically (in Word format). Upon acceptance the author will be requested to send the final version electronically (preferably in Word), accompanied by two identical print-outs of the text.
4. Authors are responsible for observing the laws of copyright when quoting or reproducing material from other sources. The copyright to articles published in *Pragmatics & Cognition* is held by the Publisher. A Copyright Assignment form will be provided to you by the Editors upon acceptance of your article. Permissions for the author to use the article elsewhere will not be withheld unreasonably upon written request.
5. Papers should be reasonably divided into sections and, if necessary, sub-sections.
6. Line drawings (FIGURES) and photographs (PLATES) should be submitted as reproducible originals, or in TIF or EPS format with a resolution of 300dpi or higher, accompanied by the original creation files and an identical hard copy. They should be numbered consecutively and appropriate captions should be provided. Reference to any Figures or Plates should be made in the main text and their desired position should be indicated on the printout.
7. TABLES should be numbered consecutively and provided with appropriate captions. They should be referred to in the main text and their desired position should be indicated on the printout.
8. QUOTATIONS should be given in double quotation marks. Quotations longer than 4 lines should be indented with one line space above and below the quoted text. All quotations should indicate their author and source (as it appears in the References list) including page reference.
9. EXAMPLES should be numbered with Arabic numerals in parentheses and set apart from the main body of the text with one line space above and below. Examples from languages other than English should be underlined (or italicized) and there should be a translation in single quotes immediately below each such example. A word-by-word gloss may be provided between the example phrase and the translation.
10. NOTES will be published as 'endnotes'. They should be kept to a minimum and should be numbered consecutively throughout the text. They should be listed in a section 'Notes' following the main text. The Notes should not contain reference material if this can be absorbed in the text and References list.
11. REFERENCES: References in the text should be as precise as possible, giving page references where necessary; for example: (Brown 1989: 252). All references in the text should appear in the References list.
12. The REFERENCES list should follow the NOTES and should start on a new page. References should be listed (1) alphabetically and (2) chronologically, beginning with the earlier ones per author. Names of Journals should be given in full with page references. Please pay special attention to the use of capitals, italics and punctuation marks given in the following examples:

    Books

    D'Agostino, F. and Jarvie, I.C. (eds). 1989. *Freedom and Rationality: Essays in Honour of John Wilkins* (Boston Studies in the Philosophy of Science 117). Dordrecht: Kluwer.

    Schauer, F. 1991. *Playing by the Rules*. Oxford: Clarendon.

    Article in book/journal

    Grice, H.P. 1975. "Logic and conversation". In P. Cole and J.L. Morgan (eds), *Syntax and Semantics, Vol. 3: Speech Acts*. New York: Academic Press, 41-58.

    Wierzbicka, A. 2000. "The semantics of human facial expressions". *Pragmatics & Cognition* 8(1): 147-183.

13. Authors are kindly requested to check their manuscripts very carefully before submission in order to avoid delays and extra costs at the proof stage. Page proofs will be sent to the (first named) author by email in PDF format. These must be returned to the journal Editor as a printout of the PDF file with corrections clearly marked, within 10 days of receipt. Any author's alterations other than typographical corrections in the page proofs may be charged to the author at the Publisher's discretion.
14. Authors of main articles will receive a free copy of the issue, plus a PDF offprint for their private use.
15. Manuscripts and all editorial correspondence should be sent to Marcelo Dascal, Editor, *Pragmatics & Cognition*, Dept. of Philosophy, Tel Aviv University, 69978 Ramat Aviv, Israel.
    Email: pragcog@post.tau.ac.il

Volume 14, No. 2 (2006)

Special Issue on Cognition and Technology
Distributed Cognition

Cognition is thinking, and thinking has been distributed for millions of years – for as long as our species has had language and tools to help us interact and collaborate and achieve far more than any of us could have done individually. But something radically new is happening to distributed cognition in recent years. With advanced technologies and especially computers and the Worldwide Web cognition can be distributed in new ways. New technologies enable new ways of interaction and distribution of cognition not only among people, but also between people and machines. This not only opens new ways to coordinate activities, but creates opportunities for new modalities of cognition to emerge. This Special Issue examines the nature of the phenomenon of distributed cognition, asks whether and how it is possible, proposes conceptual frameworks for its study and development, presents state-of-the-art technologies in this field, and looks forward at its prospects.

Contributions by Stevan Harnad and Itiel E. Dror; Pierre Poirier and Guillaume Chicoisne; John Sutton; David Kirsh; Arthur M. Glenberg; Angelo Cangelosi; Luc Steels; Todd M. Gureckis and Robert L. Goldstone; Daniel L. Schwartz and Taylor Martin; Jiajie Zhang and Vimla L. Patel; Jeffrey M. Stibel; Chris Baber, Paul Smith, James Cross, John E. Hunter and Richard McMaster; Christophe Heintz; Dimitri Kanevsky, Sara Basson, Alexander Faisman, Leonid Rachevsky, Alex Zlatsin and Sarah Conrod; Barbara Dragsted

ISSN 0929-0907